

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 29, 1999, 18:43:56 ; Search time 14.67 Seconds
(without alignments)
589.327 Million cell updates/sec

Title: US-09-077-173a-2

Perfect score: 1946

Sequence: 1 MASTESSLLRSLGLSPGPGS.....CRMAATPODSSCSTPRADRF 365

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A_Geneseq_36:*

Word size: 0

Number of hits that pass the threshold: 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1940	99.7	365	1	W23606	Human P214 recepto
2	928	47.7	375	1	R72457	Human P20 receptor
3	584	30.0	328	1	R91224	Mouse pancreas G-P
4	576	29.6	328	1	R91225	Human placenta G-P
5	571	29.3	328	1	W09433	Human placenta pur
6	510	26.2	339	1	R53752	Seven transmembran
7	510	26.2	339	1	W07617	Human G-protein th
8	510	26.2	339	1	W48733	Human R12 seven tr
9	478.5	24.6	344	1	W04247	Human G-protein co
10	449.5	23.1	370	1	W62597	Human 7-transmembr
11	423	21.7	399	1	R66932	Murine C140 recept
12	423	21.7	399	1	W01954	Murine C140 recept
13	421	21.6	395	1	R66920	Murine C140 recept
14	418	21.5	395	1	W01952	Murine C140 recept
15	403	20.7	398	1	R66921	Human C140 recept
16	403	20.7	398	1	W01953	Human C140 recept
17	397.5	20.4	425	1	R27240	Human thrombin rec
18	397.5	20.4	425	1	R60698	Fragment of the hu
19	397.5	20.4	425	1	W16314	Human thrombin rec
20	397.5	20.4	425	1	W51407	Human protease-act
21	393.5	20.2	394	1	R51408	Human protease-act
22	393	20.2	397	1	R66923	Human C140 recept
23	393	20.2	397	1	W01955	Human C140 recept
24	387	19.9	372	1	R76782	Rat delta opiate r
25	384	19.7	361	1	W70504	Leukocyte seven t
26	383	19.7	361	1	R54080	Epstein Barr virus
27	383	19.7	361	1	W53623	Human T-cell recep
28	380	19.5	333	1	R63806	Human mu opiate re
29	378.5	19.5	400	1	R71966	Human mu opiate re
30	378.5	19.5	400	1	R76780	Sequence of murine
31	377	19.4	356	1	R65188	Murine mu-subtype
32	377	19.4	356	1	R65188	Mouse delta opioi
33	377	19.4	356	1	R65188	Mouse delta opioi
34	377	19.4	356	1	R65188	Mouse delta opioi
35	376.5	19.3	398	1	R71964	Rat mu opiate rece
36	376.5	19.3	398	1	R71964	Rat mu opiate rece
37	375.5	19.3	374	1	W51406	Human mu opiate re
38	375	19.3	368	1	W54371	Human IP-10/Mig re
39	372.5	19.1	371	1	R66503	Murine delta opioi

40	372.5	19.1	334	1	W22732	Human ATP receptor
41	371.5	19.1	342	1	R25835	Human platelet act
42	371.5	19.1	398	1	W44937	Mouse mu opiate re
43	366.5	18.8	334	1	W19854	Human purinergic r
44	366	18.8	415	1	W19780	Human G-protein ch
45	365.5	18.8	424	1	W30299	Human kappa opioi

ALIGNMENTS

RESULT	ID	Description	Location/Qualifiers
1	W23606	Human P214 recepto	
2	W23606	Human P20 receptor	
3	W23606	Mouse pancreas G-P	
4	W23606	Human placenta G-P	
5	W23606	Human placenta pur	
6	W23606	Seven transmembran	
7	W23606	Human G-protein th	
8	W23606	Human R12 seven tr	
9	W23606	Human G-protein co	
10	W23606	Human 7-transmembr	
11	W23606	Murine C140 recept	
12	W23606	Murine C140 recept	
13	W23606	Murine C140 recept	
14	W23606	Murine C140 recept	
15	W23606	Human C140 recept	
16	W23606	Human C140 recept	
17	W23606	Human thrombin rec	
18	W23606	Fragment of the hu	
19	W23606	Human thrombin rec	
20	W23606	Human protease-act	
21	W23606	Human protease-act	
22	W23606	Human C140 recept	
23	W23606	Human C140 recept	
24	W23606	Rat delta opiate r	
25	W23606	Leukocyte seven t	
26	W23606	Epstein Barr virus	
27	W23606	Human T-cell recep	
28	W23606	Human mu opiate re	
29	W23606	Human mu opiate re	
30	W23606	Sequence of murine	
31	W23606	Murine mu-subtype	
32	W23606	Mouse delta opioi	
33	W23606	Mouse delta opioi	
34	W23606	Mouse delta opioi	
35	W23606	Mouse delta opioi	
36	W23606	Rat mu opiate rece	
37	W23606	Rat mu opiate rece	
38	W23606	Human mu opiate re	
39	W23606	Human IP-10/Mig re	
40	W23606	Murine delta opioi	

PS	Claim 1 Figure 1: 56bp; English
CC	This sequence represents a novel human p2 receptor. p2Y4, which has a preference for pyrimidine binding, especially uridine triphosphate.
CC	This receptor could be used to screen for novel drugs which specifically bind to it. Transgenic animals could be used to determine the physiological effects of expressing varying levels of the receptor or to identify novel agonists or antagonists. The agonists and antagonists of human p2Y4 may be used, e.g., in treatment of cystic fibrosis.
CC	Sequence 365 AA;
SO	
Query Match:	99.7%; Score 1940; DB 1; Length 365;
Best Local Similarity	100.0%; Pred. No. 2,7e-208;
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 NASTESSLLRSIGLSPGPGSEVEIDCMFDEDFKFIILPVSYAVVVLGLGNAPTLMLE 60
Dd	1 NASTSSLLRSIGLSIPGPGSSEVELDCMFDEDFKFIILPVSYAVVVLGLGNAPTLMLE 60
QY	61 IFRLRPMWDATAYMFLHALSDTLVYLSTPLTLYYYAAHNMHPGEICKFVRFLFYNNLY 120 61 IFRLRPMWDATAYMFLHALSDTLVYLSTPLTLYYYAAHNMHPGEICKFVRFLFYNNLY 120
Dd	61 IFRLRPMWDATAYMFLHALSDTLVYLSTPLTLYYYAAHNMHPGEICKFVRFLFYNNLY 120 61 IFRLRPMWDATAYMFLHALSDTLVYLSTPLTLYYYAAHNMHPGEICKFVRFLFYNNLY 120
QY	121 CSYLELTCTISVHRVYIGICHPRLALRMGRPRRLAGLICLAWLAVVAGCVNLFPPVTTSNGK 180 121 CSYLELTCTISVHRVYIGICHPRLALRMGRPRRLAGLICLAWLAVVAGCVNLFPPVTTSNGK 180
Dd	121 CSYLELTCTISVHRVYIGICHPRLALRMGRPRRLAGLICLAWLAVVAGCVNLFPPVTTSNGK 180 121 CSYLELTCTISVHRVYIGICHPRLALRMGRPRRLAGLICLAWLAVVAGCVNLFPPVTTSNGK 180
QY	181 TTVLCDDTPRPEEPFHVVHFSSAVNGLLFGVPCLVTVGYGIMARLYOPLGSAOSSSR 240 181 TTVLCDDTPRPEEPFHVVHFSSAVNGLLFGVPCLVTVGYGIMARLYOPLGSAOSSSR 240
Dd	181 TTVLCDDTPRPEEPFHVVHFSSAVNGLLFGVPCLVTVGYGIMARLYOPLGSAOSSSR 240 181 TTVLCDDTPRPEEPFHVVHFSSAVNGLLFGVPCLVTVGYGIMARLYOPLGSAOSSSR 240
QY	241 LRSRTIAVVLTVFAVCEVFPHITRTIYLAALLEADCAVLINIVNVVYVTRPPLASANC 300 241 LRSRTIAVVLTVFAVCEVFPHITRTIYLAALLEADCAVLINIVNVVYVTRPPLASANC 300
Dd	241 LRSRTIAVVLTVFAVCEVFPHITRTIYLAALLEADCAVLINIVNVVYVTRPPLASANC 300 241 LRSRTIAVVLTVFAVCEVFPHITRTIYLAALLEADCAVLINIVNVVYVTRPPLASANC 300
QY	301 LDPLYVILLTGDKRYRRQLROLCGGKRPORPTAASSIALVSLEPDSSCRMWATPODSSCSTP 360 301 LDPLYVILLTGDKRYRRQLROLCGGKRPORPTAASSIALVSLEPDSSCRMWATPODSSCSTP 360
Dd	301 LDPLYVILLTGDKRYRRQLROLCGGKRPORPTAASSIALVSLEPDSSCRMWATPODSSCSTP 360 301 LDPLYVILLTGDKRYRRQLROLCGGKRPORPTAASSIALVSLEPDSSCRMWATPODSSCSTP 360
QY	361 RADR 364 361 RADR 364
Dd	361 RADR 364 361 RADR 364
RESULT 2	
ID	R72457 standard; Protein; 375 AA. R72457
AC	R72457;
DT	29-NOV-1995 (first entry)
DE	Human P20-receptor.
KW	Epithelial mucosa; mucus; cystic fibrosis; asthma; Chronic bronchitis.
OS	Homo sapiens.
PN	W09510538.A.
PD	20-APR-1995.
PF	04-OCT-1994; U11260.
PR	15-OCT-1993; US-138137.
PA	(UMOR.) UNIV MISSOURI. (UYNC-) UNIV NORTH CAROLINA.
PI	Boucher RC, Erd LJ, Harden TK, Lustig KD, Parr CE; Pi Sullivan DM, Turner JT, Weisman GA;
DR	N-PDSB: 088134. WPI: 95-169967/22.
PT	DNA encoding human P20 receptor and null cells expressing the receptors for stimulating or inhibiting growth of cultures of mammalian cells, and for treating diseases of airway epithelial mucosa, e.g. asthma.
PS	Disclosure: page 34; 47pp; English.
CC	The sequence is that of the human P20 receptor. The receptor may be used to treat diseases of epithelial mucosal surfaces by enhancing or inhibiting mucus prodn. It may be used to treat cystic fibrosis, asthma and chronic bronchitis.

Query Match	47.7%	Score 928	DB 1	Length 375
Best Local Similarity	58.1%	Pred. No. 2,2e-95		
Matches 182	Conservative 41	Mismatches 86	Indels 4	Gaps 3

Seq	Sequence	375 AA
QY	22 EVELDQWDEDEQFKFLIPVSAAYFVYGLGNAPTLNFIIRLRPNQATATYMHIALSD 81	
Db	20 EDGRCRNEDEKLVLLPVSYGVVCLGCLNAGLVYIFLCRLTNWASTYMFHLASD 79	
QY	82 TLVYLSPTLLIYYAAAHNMHPFGTEICKFYAEFLFYMLNCSVLFLTCISVARYGICPHL 141	
Db	80 ALVYASLPLVLYYYARBGHMFSTVLCIKVAFELTYTNLCISLFLTCISVARGCVLRPL 139	
QY	142 RALNRGRRLAGLICLAVMLVYVAGCLVPNIEFTVTSNKGITVLCHDTTRPEEFDDHYHFS 201	
Db	140 RSLNRGRARVARRVAGVAVWVLYLACQAPVLYFTVTSARG-PLTCHDTSAPELFSRFVAYS 198	
QY	202 SAWVGLLEGVGCIVTIVCYGMARLTKQPLGSAQSSRL--RSRLTAVLVLYFANCFV 259	
Db	199 SVMGLLEAVFPAVILVCYVLMARLRLKPAVGISGGLPRAKRKSVRTAVVLAFAALCFE 258	
QY	260 PEHTIRITVLYARLLEADCRVLNTIVNVYKTRPLASANSCLDVLVLLTGDKYRQLRQ 319	
Db	259 PEHTYRTLYYSFRSLDSCHTLNAIMNAKTKTR-LASANSCLDVLVLYLAGQLRVRFARD 317	
QY	320 LCGGKQPQPTAA 332	
Db	318 AKPTGPSPATPA 330	
RESULT	3	
ID	R91224	standard; Protein; 328 AA.
AC	R91224;	
DT	26-AUG-1996 (first entry)	
DE	Mouse pancreas G-protein coupled receptor protein.	
KW	G-protein coupled receptor protein; G-PCR; agonist; antagonist; Cyclic fibrosis; Incontinence; diabetes; diagnosis; therapy.	
OS	Mus sp.	
PN	MO9605302-A1.	
PD	22-FEB-1996.	
PF	10-AUG-1995; J01599.	
PR	11-AUG-1994; JP-189272.	
PR	11-AUG-1994; JP-188273.	
PR	11-AUG-1994; JP-189274.	
PR	30-SEP-1994; JP-236356.	
PR	30-SEP-1994; JP-236357.	
PR	02-NOV-1994; JP-270017.	
PR	28-DEC-1994; JP-326611.	
PR	20-JAN-1995; JP-007177.	
PR	16-MAR-1995; JP-057186.	
PR	19-APR-1995; JP-093989.	
PA	(TAKE) TAKEEDA CHEM IND LTD.	
PI	Fujii R,	
PI	Fukusumi S,	
PI	Hinuma S,	
PI	Hosoya M,	
PI	Ohgi K;	
PI	Ohnaki T;	
DR	WPI: 96-139698/14.	
DR	N-PSDB: T18367.	
PT	G-protein coupled receptor protein DNA and protein - also methods for isolating (ant)agonists for treatment of cystic fibrosis,	
PT	Incontinence and diabetes	
PS	Claim 6; Page 254-55; 360pp; English.	
CC	A mouse pancreas beta-cell line MIN6-derived G-protein coupled receptor protein (G-PCR) (R91224) was identified as the	
CC	product of cDNA clone pMAH2-17 (T18367). The protein can be	
CC	obt. by expression of encoding cDNA in transformed host cells.	
CC	It was classified into a subtype within the ATP receptor, P2	
CC	purinoceptor. G-PCRs (see also R91217-25 and R91227-33) can be	
CC	used to screen agonists and antagonists that modulate G-PCR	
CC	activity, to raise antibodies and to develop assay systems.	
CC	Sequence 328 AA;	

Query Match 30.0%; Score 584; DB 1; Length 328;
 Best Local Similarity 41.6%; Pred. No. 4,5e-57;
 Matches 134; Conservative 40; Mismatches 132; Indels 16; Gaps 6;

QY 5 ESSLSLGLSPGSGSEVELDCWDEDFKILLPVSYAVVFLGGLNAPTLMLFIIRL 64
 DB 4 DNGIQAPGPP-----TTCVYRENEKQILLPPVYSVLAAGLPINICVTOICTSRALTR 60
 QY 65 RPDATATVMEHLASDTLYVLSPLTIYYAAAHNMPFGTEICKREVFLEFYNLCYVVL 124
 DB 56 RTLTRSAVYLNALADLACSLPLLIYNYAGDHMPFGDLACREVFLEFYNLHGSIL 115
 QY 125 FLTCISYHRYLGICHPRLALRW---GRPLAGLLCLAVLWVAGCLVPLNFEVTSNKG 181
 DB 116 FLTCISYHRYLGICHPRLAS--WKRGRRAAMVYGVVWLAATACLPATVPAATGIOR 173
 QY 182 TVLCHDTTREPDEEDHYHSSAVNGLLFGVPCLVLYVCGIMARL--YQPLP-GSAGSS 239
 DB 174 RTVCYDLSPLITSTRYLPYGMAITVIGFLPFIALLACYSRARRLCRODGPAGVAGER 233
 QY 240 RLRLRLTAVLYTFAVCFVPHITRTIYIARLLE-ADCRYLVNIVVYKYTRPLASAN 298
 DB 234 RSKARAAVVAFAVAFSLFEPHITKAYLAVRSTPGVSCPVLTEFAAAYKGTREFAAVN 293
 QY 299 SCLDPVLYLTGDKYRROL 320
 DB 294 SVLDPILEFTQKRRQPHDL 315

RESULT 4
 ID R91225
 AC R91225 standard; Protein; 328 AA.
 DT 26-AUG-1996 (first entry)
 DE Human placenta G-protein coupled receptor protein.
 KW G-protein coupled receptor protein; G-PCR; agonist; antagonist;
 KM cystic fibrosis; incontinence; diabetes; diagnosis; therapy.
 OS Homo sapiens.
 PN W09605302-A1.
 PD 22-FEB-1996.
 PF 10-AUG-1995; J01599.
 PR 11-AUG-1994; JP-189272.
 PR 11-AUG-1994; JP-189273.
 PR 11-AUG-1994; JP-189274.
 PR 30-SEP-1994; JP-236356.
 PR 30-SEP-1994; JP-236357.
 PR 02-NOV-1994; JP-270017.
 PR 28-DEC-1994; JP-326611.
 PR 20-JAN-1995; JP-007177.
 PR 16-MAR-1995; JP-057186.
 PR 19-APR-1995; JP-093989.
 PA (TAKE) TAKEDA CHEM IND LTD.
 PI Fujii R, Fukusumi S, Hinuma S, Hosoya M, Ohgi K;
 PI Ohnaka T;
 DR WPI; 96-139698/14.
 DR N-PSDB; T18368.
 PT G-protein coupled receptor protein DNA and protein - also methods
 PT for isolating (ant)agonists for treatment of cystic fibrosis,
 PT incontinence and diabetes
 PS Claim 6; Page 267-68; 360pp; English.
 CC A novel human placenta-derived G-protein coupled receptor protein
 CC (G-PCR) (R91225) was identified as the product of cDNA clone
 CC phM2-17 (T18368). The protein can be obtained by expression of the
 CC cDNA clone in transformed host cells. It was classified as a
 CC putinoceptor. G-PCRs (see also R91217-24 and R91227-33) can be
 CC used to screen agonists and antagonists that modulate G-PCR
 CC activity, to raise antibodies and to develop assay systems.
 SQ Sequence 328 AA;

Matches 133; Conservative 38; Mismatches 129; Indels 18; Gaps 6;

QY 10 RSLGSPGSGSEVELDCWDEDFKILLPVSYAVVFLGGLNAPTLMLFIIRL 69
 DB 9 QALGPP-----TTCVYRENEKQILLPPVYSVLAAGLPINICVTOICTSRALTR 60
 QY 70 TATVMEHLASDTLYVLSPLTIYYAAAHNMPFGTEICKREVFLEFYNLCYVFLICI 129
 DB 61 TATVTLNLAADLILACSLPLLIYNYAGDHMPFGDLACREVFLEFYNLHGSILFICI 120
 QY 130 SVHRYLGICHPRLALRW---GRPLAGLLCLAVLWVAGCLVPLNFEVTSNKGTVLCH 186
 DB 121 SFORLYGICHPRLAP--WKRGRRAAMLVYGVVWLAATACLPATVPAATGIORRTVCY 178
 QY 187 DTRPEEDHYHSSAVNGLLFGVPCLVLYVCGIMARLXQ---PLPGSAGSSRLRS 243
 DB 179 DLSPPALATHYMPGMATVIGFLPFIALLACYSCLLACRCLRODGPAPPAQ--ERRKA 237
 QY 244 LRTIAVLYTFAVCFVPHITRTIY--LARLEADCRVLNIVVYKYTRPLASANSCLD 302
 DB 238 ARMAVVAFAVAFSLFEPHITKAYLAVRSTPGVCTVLEFAAAYKGTREFAAVN 297
 QY 303 PVLYLTGDKYRROL 320
 DB 298 PLEFYTQKRRRPHDL 315

RESULT 5
 ID W09433
 AC W09433 standard; Protein; 328 AA.
 DT 01-SEP-1997 (first entry)
 DE Human placenta putinergic P-2u receptor, PNR.
 KW PNR; putinergic P-2u receptor; placental; inflammation; diagnosis;
 KW carcinoma; neoplasia; cancer; cystic fibrosis; hypertension;
 KW high blood pressure; infection.
 OS Homo sapiens.
 PN W09638558-A2.
 PD 05-DEC-1996.
 PF 03-JUN-1996; U08481.
 PR 02-JUN-1995; US-459046.
 PA (INCY-) INCYTE PHARM INC.
 PI Au-Young J, Coleman R, Stuart SG;
 DR WPI; 97-042714/04.
 DR N-PSDB; T47730.
 PT New isolated human placenta putinergic P-2u receptor (PNR) gene -
 PT used to develop prods. for the diagnosis and treatment of conditions
 PT associated with altered expression of the receptor e.g. inflammation
 PS Claim 1; Page 27-29; 36pp; English.
 CC W09433 shows a human placenta-derived putinergic P-2u receptor
 CC designated PNR. P-2u is specifically expressed in cells active in
 CC immunity. An assay for up-regulated expression of PNR can accelerate
 CC diagnosis and proper treatment of conditions caused by abnormal signal
 CC transduction due to systemic and local infections, traumatic and
 CC other tissue damage, hereditary or environmental diseases associated
 CC with hypertension, carcinomas, cystic fibrosis and other physiological
 CC or pathological problems.
 SQ Sequence 328 AA;

Query Match 29.3%; Score 571; DB 1; Length 328;
 Best Local Similarity 41.8%; Pred. No. 1,3e-55;
 Matches 133; Conservative 36; Mismatches 131; Indels 18; Gaps 6;

QY 10 RSLGSPGSGSEVELDCWDEDFKILLPVSYAVVFLGGLNAPTLMLFIIRL 69
 DB 9 QALGPP-----TTCVYRENEKQILLPPVYSVLAAGLPINICVTOICTSRALTR 60
 QY 70 TATVMEHLASDTLYVLSPLTIYYAAAHNMPFGTEICKREVFLEFYNLCYVFLICI 129
 DB 61 TATVTLNLAADLILACSLPLLIYNYAGDHMPFGDLACREVFLEFYNLHGSILFICI 120
 QY 130 SVHRYLGICHPRLALRW---GRPLAGLLCLAVLWVAGCLVPLNFEVTSNKGTVLCH 186

Db 121 SFQRLIGICHPLAF--WMKRGRRAMLVCAVWHLATVTCGLPFAITFAIGICQNRVCY 178
Oy 187 DTTREEDDHVHSSAVMGLLFGVPCVLTVLCGLMARLYQ--PLPGSNOSSRLRS 243
Db 179 DISPALATHVMPYGMALTVIGLILFPFALLACYCCLACRLCDGDAEPVAQ-ERRGKA 237
Oy 244 LRTIAVVLTVFVNCVCPHIRTITYLARLE-ADCRPLNIVNVVYKTPRLASNSCLD 302
Db 238 ARMAVVAALVGEISLPLPHIKTAYLAVRSTPGVCTVLEAFMAAYKGTREPFASNSVLD 297
Oy 303 PVLXLLTGDKYRROLROL 320
Db 298 PLTFYFTQKKRRRPHL 315
RESULT 6
ID R53752
RS3752 standard; protein; 339 AA.
AC R53752:
DT 07-FEB-1995 (first entry)
DE Seven transmembrane receptor (R12).
KW Primer: seven transmembrane receptor; receptor; amplification.
OS Homo sapiens.
FT Key
FT domain location/Qualifiers
FT 33..57
FT /label- Transmembrane domain.
FT 68..90
FT /label- Transmembrane domain.
FT 106..127
FT /label- Transmembrane domain.
FT 145..168
FT /label- Transmembrane domain.
FT 193..217
FT /label- Transmembrane domain.
FT 233..251
FT /label- Transmembrane domain.
FT 290..312
FT /label- Transmembrane domain.
FT domain
FN W09412635-A.
PN 09-JUN-1994.
PD 17-NOV-1993: UI1153.
PE 17-NOV-1992: US-977452.
PR (ICOS-) ICOS CORP.
PA Godiska R, Gray PW, Schwickart VL;
PI WPI: 94-200264/24.
DR N-PSDB: 066178.
PT DNA encoding seven transmembrane receptors - used to develop
PT prods. for use as therapeutic or diagnostic agents for conditions
PT involving the receptors.
PS Example 10: Page 81-82; 100p; English.
CC PCR using two primers (066174, 066175) was performed to amplify a
CC partial sequence of the R20 seven transmembrane coding sequence
CC (066176) from a human placenta DNA library. During the isolation
CC of the R20 gene, two weakly hybridising sequences were identified
CC which had significant homology to other seven transmembrane
CC receptors. The probe was used to screen a human genomic foetal
CC liver DNA library, and while the R20 gene could not be identified in
CC this library, several weakly hybridising clones were plaque
CC purified, subcloned and sequenced. The two clones were designated
CC R2 (066177) and R12. This is the sequence encoded by the R12 clone.
SQ Sequence 339 AA;

```
Query Match      26.2%; Score 510; DB 1; Length 339;  
Best Local Similarity 37.1%; Pred No. 8, Se. 4;  
Matches 119; Conservative 52; Mismatches 136; Indels 14; Gaps  
  
OY    27 CWFDEDFKILPVSAVVFVGGLGNAETLWLFRLRPMDATATYMHSLDLYVL 86  
      | | | | | | | | | | | | | | | | | | | | | |  
Db     23 CGQEPLEMLTASTYLDEFILALVTNTALTLMFLINDHSGSPANVFMHLNVADISCVL 82
```

QY	87	SLEPTLYAAHNMPEGTETICKVREFLPYMNLYCSVLPTGCSVNHXYGDCIPRLRLM	146
		: : : : : : : : : : : :	
Db	83	VLPRTLVHSEGNMPEGLACIRLTGLTFELTMNASTIYFTLCISADPELIVHPVSLK	1422
QY	147	GRPLRAGLCLAWLVVAGCLVPLFEVTSNKGTVLCHDTPRPEEDVYHFFSSAVMG	206
		: : : : : : : : : : : : : :	
Db	143	RRLPLAHLCAPFLMVVAVYAMAPLLVSPQVYQHTHYVCQLRL- EKASHIALSLAV--	200
QY	207	LLFEGVCLVTLVCCYGLMARLYOPLPGSAOSSRLRSRLRTIAVLTVFAVCFFPHITRT	266
		: : : : : : : : : : : : : :	
Db	200	-AFEPFTVTVTCYLLIRSLRQGL--RVKRLKTKAVRIALVLAFLVCFPYHVNRS	256
QY	267	IYYL-ARLEBDCVNLIVNVYKVTBPLASANSCLDPVLYLTGDKYRQL-ROLG--	3233
		: : : : : : : : : : : : : :	
Db	257	VYVLVHRSHGASCATORILANRITSLGNSLNGALDPIMYFVAEKFHALCNLGGKR	316
QY	323	--GGRF--QPRTAASSLALVS 339	
		: : : : : : :	
Db	317	LKGPPSEGRKTNESSLSAKS 337	
		: : : : : : :	
RESULT	7		
ID	W07617	standard; Protein; 339 AA.	
AC	W07617		
DT	28-FEB-1997	(first entry)	
DE	Human G-protein thrombin-like receptor.		
KW	G-protein; thrombin; receptor; diagnosis; therapy; agonist;		
KW	antagonist; haemophilia; wound healing; restenosis; angina;		
KW	inflammation.		
OS	Homo sapiens.		
PN	MO6363438-A1.		
PD	12-DEC-1996.		
PF	06-JUN-1995; U07180.		
PR	06-JUN-1995; WO-007180.		
PA	(HUMA-1) HUMAN GENOME SCI INC.		
PI	Gocayne JD, Li Y, Ruben SM;		
DR	WPI: 97-043073/04.		
DR	N-PSDB: T44092.		
PT	Human G-protein thrombin receptor, HIBEB69 - useful to identify		
PT	(ant.)agonists, for treatment of angina, restenosis, wound healing		
PT	etc.		
PS	Claim 1; Fig 1a-D; 58pp; English.		
CC	A novel human 7-transmembrane receptor (W07617) has been putatively		
CC	identified as a G-protein thrombin receptor. Its amino acid		
CC	sequence was deduced from a cDNA clone (T44092) discovered in a		
CC	human infant brain cDNA library. Recombinant receptor can be		
CC	produced in host cells (e.g. E. coli, COS-7, Sf9) and used to		
CC	raise antibodies or to screen for (ant.)agonist cpds. Agonists		
CC	can be used to treat e.g. haemophilia or to promote wound		
CC	healing; antagonists can be used to treat inflammation, angina,		
CC	restenosis, stroke, etc.		
QD	Sequence 339 AA;		

Query Match	26.28	Score 510	DB 1	Length 339
Best Local Similarity	37.18	Pred. NO. 8.5e-49		
Matches 119	Conservative 52	Mismatches 136	Indels 14	Gaps 7

QY	27	CWDEDEFFILPVSAYAVVYGLGNAPTLMLEIFLRPMDATATYFHALSDTLTYVL	86
		- : : : : : - : : : : : - : : : : : - : : : : : - : : : : : -	
Db	23	CGQETPLENMFAFYLDFDLFTLAVGNTLALMLERIDHKSSTPRAWFMHLAAVADSLCV	82
		- : : : : : - : : : : : - : : : : : - : : : : : - : : : : : -	
QY	87	SLPTLIYYAAHNMPFETGECIKFVRFETVYNTLCSVYFLCISVHYRILGICHPDRALRW	146
		- : : : : : - : : : : : - : : : : : - : : : : : - : : : : : -	
Db	83	VLPRLYVHFGSNHMPGEICACRLTGLGFELNMYASTIFELCISADRLALVHPKSKL	142
		- : : : : : - : : : : : - : : : : : - : : : : : - : : : : : -	
QY	147	GRPLAGLICLAVNLVYAAGCLVPLNFETVYNSKGTIYVCHPTTRPEEDHYVHPSAAYMG	206
		- : : : : : - : : : : : - : : : : : - : : : : : - : : : : : -	
Db	143	RRLPLAHACAFNLVYAVAAAPLILVSEQTQVNTHTVYCDLQYR-EKASHALVSLAA--	200
		- : : : : : - : : : : : - : : : : : - : : : : : - : : : : : -	
QY	207	LLEGVPCLLVYCVGLMAARLQDPLPSAAGSSRLSLRIAYVLVYAVGVFPRHITRL	266
		- : : : : : - : : : : : - : : : : : - : : : : : - : : : : : -	

Db 200 -AFTEPFITVTCYLLIIRSLRQGL--RVEKRLKTKAVMAIVLAIFLCVCPYHVNS 256
 QY 267 IYVL-ARLEADCRVLNVVYKYTRPLASANSCLDPVLYLTGDKYRQL-ROLCG-- 323
 Db 257 VYLVHRSHGASCATQRIALANRITSCLTSLNGALDPIIMYFVAKFKHALCNLCGR 316
 QY 323 --GGRP--QPRTAASSLALVS 339
 Db 317 LKGPSPSEFGKTNESLSAKS 337

RESULT 8

W48733
 ID W48733 standard; Protein: 339 AA.
 AC W48733;

DT 25-SEP-1998 (first entry)
 DE Human R12 seven transmembrane (7TM) receptor.
 KW V28; placenta; seven transmembrane receptor; 7TM; signal transduction;
 OS Homo sapiens.

FT Key Location/Qualifiers

FT Domain 33..57
 FT Domain /note= "Transmembrane domain 1"
 FT Domain /note= "Transmembrane domain 2"
 FT Domain /note= "Transmembrane domain 3"
 FT Domain /note= "Transmembrane domain 4"
 FT Domain /note= "Transmembrane domain 5"
 FT Domain /note= "Transmembrane domain 6"
 FT Domain /note= "Transmembrane domain 7"

PN US755804-A.

PD 02-JUN-1998.

PE 17-NOV-1993; 153848.

PR 17-NOV-1992; US-977452.

PA (ICOS-) ICOS CORP.

PI Godtska R. Gray PW, Schweickart VL;

DR WPI; 98-332132/29.

DR N-PSD; V18356.

PT DNA encoding V28 seven transmembrane receptor polypeptide - useful

PT for producing recombinant polypeptide and anti-V28 antibodies, and

PT in screening assays for V28 agonists and antagonists

PS Example 10; Columns 87-90; 56pp; English.

CC The present sequence represents the human R12 protein which is encoded

CC by the full length genomic R12 clone (V18356) which was isolated from

CC a human genomic fetal liver DNA library. The invention claims for the

CC full length V28 genomic DNA sequence (V18343) isolated from a human

CC placenta genomic library. The V28 (W48722) and R12 proteins are seven

CC transmembrane (7TM) receptors which are probably involved in signal

CC transduction. The invention also claims that cells transformed with V28

CC DNA can be used to produce the recombinant polypeptide, to produce

CC anti-V28 antibodies or in screening assays for V28 agonists or

CC antagonists. The antibodies, agonists and antagonists could then be

CC used to modulate V28 receptor-ligand binding, for e.g. in immunological

CC and/or inflammatory events in vivo.

CC Sequence 339 AA;

SQ

Query Match 26.2%; Score 510; DB 1; Length 339;

Best Local Similarity 37.1%; Pred. No. 8.5e-49;

Matches 119; Conservative 52; Mismatches 136; Indels 14; Gaps 7;

QY 27 CWFDEDFKFIILPVSYAVYVGLGNAPTLMLFRLRPMNATATYMHLSLDTLYL 86
 Db 23 CGGEFPLEMLFASYLLDFILALVGNITLALWLFIDHSGRPANVLAHLVADISCL 82
 QY 87 SLPTLIYYAAHNNHPFGTEICKFVRFYVNLVYCSVFLTCISVHRVYIGICHPRLALRW 146
 Db 83 VLPRLVYVHFSGNHNPFGIACRLTGLFELVLMNMAISYITLTICISADRLAIYHPVKSXL 142

QY 147 GRPRLAGLCLAVWLVVAGCLVPLNFEVTTSNKGTTLCHDTPREEDHYHSSAVNG 206
 Db 143 RRPVLAHLACALVWLVVAVNAPPLVSPOTVOTNHTVYLOLYR-EKASHHALVSLAV-- 200
 QY 207 LLEGVPCVTLVYCYGLMARRLYQPLPGSAOSSRSLRSLTIAVLTVAFCVPEPHITRT 266
 Db 200 -AFTEPFITVTCYLLIIRSLRQGL--RVEKRLKTKAVMAIVLAIFLCVCPYHVNS 256
 QY 267 IYVL-ARLEADCRVLNVVYKYTRPLASANSCLDPVLYLTGDKYRQL-ROLCG-- 323
 Db 257 VYLVHRSHGASCATQRIALANRITSCLTSLNGALDPIIMYFVAKFKHALCNLCGR 316
 QY 323 --GGRP--QPRTAASSLALVS 339
 Db 317 LKGPSPSEFGKTNESLSAKS 337

RESULT 9

W04247
 ID W04247 standard; Protein: 344 AA.

AC W04247;

DT 13-DEC-1996 (first entry)

DE Human G-protein coupled receptor GPR4.

KW G-protein coupled receptor; GPR4; signal transduction; agonist;

KW antagonist; cell proliferation; cancer; tumour; asthma; allergy;

KW diagnosis.

OS Homo sapiens.

PN W09630406-AI.

PD 03-OCT-1996.

PE 30-MAR-1995; U04079.

PR 30-MAR-1995; WO-004079.

PA (HMDA-) HUMAN GENOME SCI INC.

PI Bult C, Cao L, Gents R, Li Y, Nij J, Rosen CA;

PI Sutton G;

DR WPI; 96-455278/45.

DR N-PSD; T33905.

PT Nucleic acid encoding human G-protein coupled receptor - for

PT diagnosing diseases, and identifying (ant)agonists for asthma and

PT allergy treatment

PS Claim 1; Page 56-57; 87pp; English.

CC Novel human G-protein coupled receptor GPR4 (W04247) was identified

CC as the product of a cDNA clone (T33905) isolated from a human 12-wk

CC embryo. The protein is also expressed in the human heart, spleen

CC and leukocytes. It shows 82% identity over a 291-amino acid

CC stretch with a chick orphan G-protein coupled receptor. Potential

CC ligands include thrombin, chemokine and platelet activating

CC factor. Recombinant GPR4, GPR1, GPR2 and GPR3 (see also W04244-46)

CC can be expressed in e.g. E. coli, COS or insect cell hosts for use in

CC identifying (ant)agonist cpts. Agonists may be used to treat asthma,

CC Parkinson's disease, hypotension, osteoporosis etc., and antagonists

CC to treat ulcers, asthma, allergies, etc.

CC Sequence 344 AA;

SQ

Query Match 24.6%; Score 478.5; DB 1; Length 344;

Best Local Similarity 33.4%; Pred. No. 2.8e-45;

Matches 99; Conservative 66; Mismatches 126; Indels 5; Gaps 4;

QY 27 CWFDEDFKFIILPVSYAVYVGLGNAPTLMLFRLRPMNATATYMHLSLDTLYL 86
 Db 9 CFYNDSEFKTLIGCMFNMVYVGLISNCVAITFICVLRNRETTYTMNLMASDLLEVF 68
 QY 87 SLPTLIYYAAHNNHPFGTEICKFVRFYVNLVYCSVFLTCISVHRVYIGICHPRLALRW 146
 Db 69 TLPRILFFITRN-WPFGDLCKISVLMFLTYMYSIILFLTICISDRFLATYPPKSKTL 127
 QY 147 GRPRLAGLCLAVWLVVAGCLVPLNFEVTTSNKGTTLCHDTPREEDHYHSSAV 206
 Db 128 RTRNAKIVCTGSMVLTIVGSAVAVQSTHSGNNAADACENPEATKWTYLSRIYF 187
 QY 205 MGLL-FGVPCVTLVYCYGLMARRLYQPLPGSAOSSRSLRSLTIAVLTVAFCVPEPHI 263

Db 188 IEIVGFIPILNLTSCSMWKLITKRPVLTSRSKINKTKVLMFVHLIFGCFVPEYNI 247
QY 264 TRITTYLABLE-ADCRVLNINVVYKTRPLASANSCLDPVLLTGDKYRQLR 318
Db 248 NLILYSLVROTFNCSVAAVRMTYPTICIAVSNCCFPIYVFTSDTIONSIX 303

RESULT 10
ID W62597 standard; Protein: 370 AA.
AC W62597;
DE 07-OCT-1998 (first entry)
KW Human 7-transmembrane receptor, HB8CS41.
KW G-protein coupled receptor; HB8CS41; treatment; diagnosis; infection;
KW HIV-1; HIV-2; cancer; screening; human; 7-transmembrane receptor;
KW Parkinson's disease; vaccine; anorexia; bulimia; hypertension; asthma;
KW urinary retention; osteoporosis; ulcer; allergy; neurological disorder.
OS Homo sapiens.
PN EP-853126-A2.
PD 15-JUL-1998.
PR 08-JAN-1998; 300130.
PR 10-JAN-1997; US-781250.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Bergsma DJ, Mao JY, Sahe G, Van HORN S;
DR N-PSDB; V38565.
PT Human G-protein coupled receptor, HB8CS41 - useful e.g. in treatment
PT screening of antagonistic or agonistic compounds
PS Claim 15; Fig 1; 20pp; English.
CC This represents a human 7-transmembrane receptor, HB8CS41. HB8CS41 is a
CC human G-protein coupled receptor and can be used in vaccines to protect
CC mammals from infections e.g. bacterial, fungal, protozoan and viral
CC infections (especially infections caused by HIV-1 or HIV-2), cancers and
CC Parkinson's disease. They can also be used in the treatment of anorexia,
CC bulimia, hypertension, asthma, urinary retention, osteoporosis, ulcers,
CC allergy and neurological disorders. The HB8CS41 polypeptides can be used
CC to produce antibodies which are used to treat such infections, isolate
CC or identify clones expressing the polypeptide or to purify the
CC polypeptide. The polypeptides can also be used to screen for compounds
CC binding to, activating or inhibiting activation of HB8CS41. They can be
CC used to diagnose diseases and susceptibility to diseases related to
CC expression or activity of HB8CS41.
SQ Sequence 370 AA;

Query Match 23.1%; Score 449.5; DB 1; Length 370;
Best Local Similarity 33.4%; Pred. No. 5.4e-42;
Matches 102; Conservative 67; Mismatches 131; Indels 5; Gaps 5;

QY 14 LSPGPGSEVELDQWEDFKFILLPVSYAVFVLGGLNAPLTLMTFLRLPMDATATY 73
Db 18 LRLRLGATNANTCIVDSFKYNGAVSVFLGLITNSVSLFCFRMKRSETAIF 77
QY 74 MFLALSDTYLVSLPTLIYYAAHNMPTGETCKFVRELVNKLCSVFLTCISVHR 133
Db 78 ITNLASDILFVCTLPKFT-FYNENRMPFGDICKSGAFLTNIGSMFLTCISVDR 136
QY 134 YLQICHLRLRWGRPLACLCLAVN-LVYAGCLVNLFEVTTNSNGTVLCHDTRPE 192
Db 137 FLAIVPFRSRTIRTRNSAIVCGWILVLSGISASLSTTNVNAITT-CEGFSKR 195
QY 193 EFDHYHFFSSAVNGL-FGYPCLVTLVYGLMARLYQPLPGSAOSSRLRSRTIAVL 251
Db 196 VMKTYLSKRIFFLEVGFILPLINAVSCSSVYLTLLKRPATLSQIGNKKKVLMTIVHM 255
QY 252 TVFAVCFPHIRITTYLABLE-ADCRVLNINVVYKTRPLASANSCLDPVLLTGL 310
Db 256 AVTVVCCVPIVNSVLELALVRSQAITNCFERPAKIMPTITLCLATINCCFDPPIYFTL 315
QY 311 DKYR 315
Db 316 ESFOK 320

RESULT 11
ID R66922 standard; Protein: 399 AA.
AC R66922;
DE 22-AUG-1995 (first entry)
KW Murine C140 receptor deduced from cDNA.
KW G-protein-coupled receptor; G-protein; C140 receptor.
OS Mus musculus.

PN W09503318-A.
PD 02-FEB-1995.
PR 26-JUL-1994; U08536.
PR 26-JUL-1993; US-097938.
PA (CORR-) COR THRAPEUTICS.
PI Scarborough RM, Sundell J;
DR WPI; 95-075182/10.
DR N-PSDB; Q84559.
PT New DNA encoding recombinant C140 receptor - and novel agonists
PT and antagonists and specific antibodies with therapeutic and
PT diagnostic applications.
PS Example; Fig 10; 57pp; English.
CC A cDNA library from a mouse stomach was constructed in lambda gt10
CC and screened with a probe encompassing the C140 genomic clone (see
CC Q84557). A single phage clone was isolated and cut with EcoRI. The
CC insert was cloned in pBluescript and pGS5 and sequenced. The
CC complete nt sequence and deduced AA sequence is given in Q84559 &
CC R66922. 5' RACE resulted in the addition of only 27 bps to the 5'
CC end. The 5' end of the apparent coding region differs from the 5'
CC end of the ORF of genomic DNA; it is believed that the 5' end of
CC the cDNA sequence is correct.
SQ Sequence 399 AA;

Query Match 21.7%; Score 423; DB 1; Length 399;
Best Local Similarity 31.8%; Pred. No. 5.3e-39;
Matches 107; Conservative 55; Mismatches 141; Indels 34; Gaps 10;

QY 36 ILPVSYAVFVLGGLNAPLTLMTFLRLPMDATATYFHLASDLYVLSPTLIYY 95
Db 78 VFLPVYIIVFVLGSLNSGALWIFLFRKKRPAYVYANLALDLSVYMPKISYH 137
QY 96 AAMNHPFGEICKFRFLFYMLYCSVFLTCISVHRVIGICHPALRMGRPR---L 151
Db 138 LHGNMVGELCKVLIGFYGMYCSILFMTLSVORWVYNPM---GHPRRKANI 192
QY 152 AGILCLAVMLVAGCLVP-----NLFEVTTNSNGTVLCHDTRPEEF---DHYHSSA 203
Db 193 AVGSLAIVMLITLVITPLVYKOTIYIPALNITT---CHDV-PESVLVGDMFNFTSL 248
QY 204 VMGLFVPCLVTLVYGLMARLYQPLPGSAOSSRLRSRTIAVLTVFAVCFVPHI 263
Db 249 AIG-VFLPFLTLASVYVLMIKTLRSSAMDESEKRRQRAIRLITVLAIVFCFAPSNL 307
QY 264 TRITTYLABLEDCRYLVNINVVYKTRPLASANSCLDPVLLTGDKYRQLRO--LC 321
Db 308 LLYVHYELIKTORSHYV---ALYVLAICLSTLNSCIDPFIYVFSKDFDARNALLC 363
QY 322 GGGKPPRTAASSLALVSLPDSRSSCHMAATPPDSSCS 358
Db 364 RSVRTVNRMO-----ISLSSNFKSRKSGSYSSSTS 394

RESULT 12
ID W01954 standard; Protein: 399 AA.
AC W01954;
DE 02-APR-1997 (first entry)
KW Murine C140 receptor.
KW C140 receptor; G-protein linked; seven pass; agonist;
KW antagonist; hypertension; hypotension; blood pressure.
OS Mus sp.
FH key Location/Qualifiers

```

FT      peptide
FT      1..31
FT      /note="the signal peptide differs from that
FT      encoded by a genomic DNA sequence for
FT      this receptor (see W01952), the signal
FT      sequence given here is believed to be
FT      the correct sequence"
FT      32..399
FT      protein
FT      /note="mature protein"
FT      W09623225-A1.
FT      PD 01-AUG-1996.
FT      PF 25-JAN-1996; U01179.
FT      PR 25-JAN-1995; US-390301.
FT      PA (CORT-) COR THERAPEUTICS INC.
FT      P1 Scarborough NM, Sundelin J;
FT      DR WPI, 96-362813/36.
FT      N-PSDB: T32038.
FT      PT Vector for expression C140 cell surface receptor in host cell -
FT      useful to identify C140 agonist and antagonists, which are
FT      antihypertensives and elevators of blood pressure, respectively
FT      PS Example 4; Fig 10A-B; 60BP; English.
FT      CC W01954 represents the murine C140 receptor (C140R). DNA encoding C140R
FT      CC may be engineered so as to allow the recombinant expression of C140R in
FT      CC a suitable host cell, i.e. by removing the native expression-control
FT      CC sequences and replacing them with control sequences operable in the host.
FT      CC Such a recombinant receptor can be expressed on the surface of oocytes,
FT      CC this provides a good assay system for identifying agonists/antagonists
FT      CC of C140R. The C140 receptor is a G-protein linked receptor and a member
FT      CC of the "seven-pass" transmembrane receptor superfamily (peptide chain
FT      CC of the receptor passes through the cell membrane seven times, producing
FT      CC seven transmembrane regions within the receptor molecule). The C140
FT      CC receptor is involved in controlling blood pressure. C140 antagonists
FT      CC (see W01942-W01951) are useful to inhibit signalling from this
FT      CC receptor, resulting in an increase in blood pressure and are therefore
FT      CC useful in pharmaceuticals for the treatment of hypotension (low blood
FT      CC pressure). Conversely agonists (see W01914-W01941) of C140 are useful
FT      CC in pharmaceuticals for the treatment of hypertension (high blood
FT      CC pressure).
FT      Sequence 399 AA:

```

DT	22-AUG-1995 (first entry)
DE	Murine C140 receptor.
KM	G-protein-coupled receptor; G-protein; C140 receptor.
OS	Mus musculus.
FH	Key
FT	protein
FT	/label= tentative signal sequence
FT	78..100
FT	/label= transmembrane I
FT	108..128
FT	/label= transmembrane II
FT	168..191
FT	/label= transmembrane III
FT	168..191
FT	/label= transmembrane IV
FT	204..224
FT	/label= transmembrane V
FT	266..286
FT	/label= transmembrane VI
FT	304..325
FT	/label= transmembrane VII
FT	29
FT	/label= Asn linked glycosylation site
FT	34..35
FT	/label= protease receptor cleavage
FT	200
FT	/label= Asn linked glycosylation site
PM	M09503318-A.
PD	02-FEB-1993.
PF	26-JUL-1994; U08536.
PR	26-JUL-1993; US-097938.
PA	(CORF-) COR THERAPEUTICS.
PI	Scarborough RM, Sundelin J,
DR	Wpi: 95-075182/10.
DR	N-PSSB; Q84557.
PT	New DNA encoding recombinant C140 receptor - and novel agonists
PT	and antagonists and specific antibodies with therapeutic and
PT	diagnostic applications.
PS	Disclosure; Fig 1: 57pp; English.
CC	A mouse cosmid genomic library (obtd. from Dr R.A Wetzel, Washington
CC	Univ. School of Medicine, St Louis, Missouri) was screened with two
CC	32P-labeled oligos corresp. to bp 190-249 and 742-801 of the bovine
CC	substance K receptor cDNA. In one of the clones isolated (C140) the
CC	hybridising region was localised to a 3.7 kb psti fragment. This
CC	fragment was subcloned into pBluescript vector. The hybridising and
CC	adjacent regions were sequenced. The nt sequence and the deduced AA
SO	Sequence are given in Q84557 & R69920 respectively.
	Sequence 395 AA;

PT and antagonists and specific antibodies with therapeutic and
 PT diagnostic applications.
 PS Disclosure; Fig 2; 57pp; English.
 CC The availability of genomic DNA encoding the mouse protease
 CC C140 receptor (see 084557) permitted the retrieval of the corresp.
 CC human gene. A human genomic library cloned in the vector EMBL3 was
 CC screened using the entire coding region of the murine clone as a
 CC probe. The recovered human gene including the DNA sequence and the
 CC deduced AA sequence are shown in 084558 & R66921. Subsequent
 CC experiments indicated that the human C140 gene is located in the
 CC same region of the long arm of chromosome number 5 (5q12-5q13)
 CC as has been reported for the human thrombin receptor gene.
 SQ Sequence 398 AA;

Query Match 20.7%; Score 403; DB 1; Length 398;
 Best Local Similarity 32.1%; Pred. No. 9,1e-37;
 Matches 95; Conservative 57; Mismatches 124; Indels 20; Gaps 9;
 QY 36 ILLEPSYAVVFLGIGLNAPTLMTLFIPLRPMDATATYMFHLSDTLYVSLPTLIYY 95
 Db 77 VFLPIVYTVFVGLPSNGMALWFLFRTKKHRAVITMANLADLSVIMFPLKIAIH 136
 QY 96 AAHNHMPGTEICKFVRFLFYWNLYCSVLFITCISVHRTIGICHPRLALRMGRPRLAGIL 155
 Db 137 IHGNNMITYGEALCVLIGFEGNMYCSILFMTCLSVQRVWIVNMGHSR-KKANIALGI 195
 QY 156 CLAVWLVVAGLVP-----NLEFYTTSNKGTVVLCSDTTRPEEF---DHYVHSSAVMGL 207
 Db 196 SLAIWLLILVLTIPLYVVKOTIFIPALNITF--CHDVL-PEQLVGDMPFYFLSLAIG- 251
 QY 208 LFGVPCLVTVLCYGMARLRYQPLPGSAOSSRLSLRTIAVLVLEAVCFVPEHITRTI 267
 Db 251 VFLEPAFLTASAYVLMRLMRSSANDENSEKKRRAIKLIVTILAMYLICFTPSMLLV 310
 QY 268 YVLAARLLEADCRVLINIVNVYKVRPLASANSCLDPVLYLLTGDKYRRLRQ--LC 321
 Db 311 HYF--LIKSGOQ--SHYVALYIVALCLSTLNSCIDPEFYVYFVSHDFRDNANALLC 362

Search completed: December 29, 1999, 19:43:39
 Job time: 3583 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 07:20:25 : Search time 104.56 Seconds
(without alignments)
83.748 Million cell updates/sec

Title: US-09-077-173a-3

Perfect score: 35
Sequence: 1 CAGATCTAGATGATGTTCTACACTCTTACGTC 35

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Word size: 0

Number of hits that pass the threshold: 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20	57.1	1209	1	T39519	Flea calreticulin
C 2	20	57.1	1589	1	T39516	Flea calreticulin
C 3	20	57.1	1209	1	T39517	Flea calreticulin
C 4	20	57.1	1589	1	T39518	Flea calreticulin
C 5	19.4	55.4	110000	1	X20248_08	Continuation (9 of Human secreted pro
C 6	19.4	54.3	1291	1	V59636	Human secreted pro
C 7	18.6	53.1	5503	1	V18187	Fanconi anaemia of
C 8	18.6	53.1	9183	1	V60751	HIV-1 strain YBR30
C 9	18.6	53.1	1539	1	V60753	HIV-1 strain YBR30
C 10	18.4	52.6	1990	1	Q33064	Encodes Babesia bo
C 11	18.4	52.6	1990	1	T18995	Babesia merzoeite
C 12	18	51.4	846	1	X37420	Human secreted pro
C 13	17.8	50.9	150	1	X12163	Human biallelic pro
C 14	17.6	50.3	1192	1	T08635	HPV E6-binding pro
C 15	17.6	50.3	97	1	T23922	Human gene signatu
C 16	17.6	50.3	12047	1	V17618	Homo sapiens gland
C 17	17.6	50.3	12047	1	X24752	Human glandular ka
C 18	17.4	49.7	10095	1	O81762	Japanese Black pin
C 19	17.4	49.7	2582	1	X20558	Polynucleotide seq
C 20	17.2	49.1	1350	1	Q54397	Feline immunodef
C 21	17.2	49.1	1230	1	Q55280	Sequence of the re
C 22	17.2	49.1	1418	1	Q85885	Open reading frame
C 23	17.2	49.1	150	1	O85886	Stimulocalcin alph
C 24	17.2	49.1	892	1	T18979	Prostate-specific
C 25	17.2	49.1	2125	1	T36787	FIV gag gene. Reco
C 26	17.2	49.1	1353	1	T47239	FIV gag gene 3' PC
C 27	17.2	49.1	23	1	T47249	Feline immunodef
C 28	17.2	49.1	1237	1	T74084	Feline immunodef
C 29	17.2	49.1	1353	1	T49076	Feline immunodef
C 30	17.2	49.1	9472	1	V25601	Feline immunodef
C 31	17.2	49.1	4616	1	V26975	Consensus VCP255 H
C 32	17.2	49.1	2032	1	V26971	Feline immunodef
C 33	17.2	49.1	2032	1	V26974	Consensus I3L prom
C 34	17.2	49.1	4150	1	V26976	Consensus VCP329 H
C 35	17.2	49.1	237326	1	V57903	Hereditary haemoch
C 36	17.2	49.1	235033	1	V57926	FIV F14 clone. Pro
C 37	17.2	49.1	9474	1	V58053	FIV-NCSU1 clone JS
C 38	17.2	49.1	9474	1	V57355	FIV-NCSU1 clone JS
C 39	17.2	49.1	9471	1	V57356	FIV-NCSU1 clone JS

C 40	17	48.6	10723	1	012787	Dengue 2 virus gen
C 41	17	48.6	1349	1	057701	K.lactis transalido
C 42	17	48.6	1226	1	057702	K.lactis transalido
C 43	17	48.6	4090	1	085027	Entamoeba histolyt
C 44	17	48.6	133894	1	T13635	ACNPV genomic DNA
C 45	17	48.6	423	1	T62727	DNA from construct

ALIGNMENTS

RESULT 1	
TD T39519/c	
AC T39519	standard; cDNA; 1209 BP.
DT 12-DEC-1996	(first entry)
DE Flea calreticulin cDNA clone nctCal1209 (complement).	
KW Calreticulin; flea; haematophagous insect; allergic dermatitis;	
KW vaccine; therapy; nctCal1209; ss.	
OS Ctenocephalides felis.	
PN WO9628469-A1.	
PD 19-SEP-1996.	
PF 08-MAR-1996; U03133.	
PR 09-MAR-1995; US-401509.	
PA (HESK-) HESKA CORP.	
PI Rushlow KE, Stiegler GL;	
DR WPI; 96-442861/44.	
PT Haematophagous insect calreticulin protein - used to reduce insect	
PS infestation and desensitise patients to allergic dermatitis	
PS Claim 16: Page 72; 86pp; English.	
CC The antisense strand (T39519) of nucleic acid nctCal1209 (T39517)	
CC hybridises under stringent conditions to a flea calreticulin gene.	
CC nctCal1209 codes for a flea calreticulin protein (W04171) that can	
CC be used to alter the blood feeding behaviour of haematophagous	
CC insects or to desensitise an animal to allergic dermatitis. The	
CC antisense strand, when administered to an animal, reduces	
CC calreticulin activity in insects feeding on the animal, thereby	
CC reducing the insect burden on the animal and in the environment	
CC of the animal.	
SQ Sequence 1209 BP; 292 A; 280 C; 212 G; 425 T;	

Query Match	57.1%; Score 20; DB 1; Length 1209;
Best Local Similarity	82.1%; Pred. No. 5.7;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	
QY 1 CAGATCTAGATGATGTTCTTACACTCT 28	
DB 1007 CAGGTCAAGATGCTAGTCTACGCTCT 980	
RESULT 2	
TD T39516	standard; cDNA; 1589 BP.
AC T39516	
DT 12-DEC-1996	(first entry)
DE Flea calreticulin cDNA clone nctCal1589	
KW Calreticulin; flea; haematophagous insect; allergic dermatitis;	
KW vaccine; therapy; nctCal1589; ss.	
OS Ctenocephalides felis.	
PN WO9628469-A1.	
PD 19-SEP-1996.	
PF 08-MAR-1996; U03133.	
PR 09-MAR-1995; US-401509.	
PA (HESK-) HESKA CORP.	
PI Rushlow KE, Stiegler GL;	
DR WPI; 96-442861/44.	
DR P-PSDB; W04171.	
PT Haematophagous insect calreticulin protein - used to reduce insect	
PT infestation and desensitise patients to allergic dermatitis	

PS Claim 7: Page 66-68: 86bp: English.
CC A CDNA clone (T39516), designated nctCal1589, codes for a flea
CC calreticulin protein (W04171) that can be used to alter the blood
CC feeding behaviour of haematophagous insects or to desensitise an
CC animal to allergic dermatitis. Its sequence was deduced from
CC clones nctCal1655, nctCal1750 and nctCal1218 isolated from a
CC ctenocephalides felis salivary gland cDNA library by PCR
CC amplification using flea calreticulin gene-specific primers (see
CC also T39529-32). nctCal1589 and other calreticulin nucleic acids
CC (see also T39515 and T39517-20) can be used to produce recombinant
CC calreticulins, or can be administered to an animal to protect it
CC against haematophagous insect infestation.
SQ Sequence 1389 BP: 352 A: 266 C: 335 G: 436 T:

Query Match 57.1%; Score 20; DB 1: Length 1589;
Best Local Similarity 82.1%; Pred. No. 6;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 CAGATCTAGACTAGTCTTCTACACTCT 28
DB 353 CAGCTCAAGATGCTAGGTTCTACGCTCT 380

RESULT 3
T39517
ID T39517 standard: cDNA: 1209 BP.
AC T39517:
DT 12-DEC-1996 (first entry)
DE Flea calreticulin cDNA clone nctCal1209.
KW Calreticulin; flea; haematophagous insect; allergic dermatitis;
KW vaccine; therapy; nctCal1209; ss.
OS Ctenocephalides felis.
PN MO9628469-A1.
PD 19-SEP-1996.
PR 08-MAR-1996; U03133.
PR 09-MAR-1995; US-401509.
PA (HESK-) HESKA CORP.
PI Rushlow KE, Stiegler GL;
DR WPI: 96-442861/44.
DR P-PSDB; W04171.
PT Haematophagous insect calreticulin protein - used to reduce insect
PT infestation and desensitise patients to allergic dermatitis
PS Claim 7: Page 71: 86bp: English.
CC Nucleic acid nctCal1209 (T39517) comprises the coding region (minus
CC the stop codon) encoding flea calreticulin PctCal403 (W04171), a
CC protein that can be used to alter the blood feeding behaviour of
CC haematophagous insects or to desensitise an animal to allergic
CC dermatitis. The coding sequence was deduced from a cDNA clone
CC (T39516) isolated from a cDNA library prep. from the salivary
CC glands of Ctenocephalides felis. Calreticulin nucleic acids (see
CC also T39515 and T39518-20) can be used to produce recombinant
CC calreticulins, or can be administered to an animal to protect it
CC against haematophagous insect infestation.
SQ Sequence 1209 BP: 425 A: 212 C: 280 G: 292 T:

Query Match 57.1%; Score 20; DB 1: Length 1209;
Best Local Similarity 82.1%; Pred. No. 5.7;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 CAGATCTAGACTAGTCTTCTACACTCT 28
DB 203 CAGCTCAAGATGCTAGGTTCTACGCTCT 230

RESULT 4
T39518/c
ID T39518 standard: cDNA: 1589 BP.
AC T39518:
DT 12-DEC-1996 (first entry)
DE Flea calreticulin cDNA clone nctCal1589 (complement).
KW Calreticulin; flea; haematophagous insect; allergic dermatitis;

KW vaccine; therapy; nctCal1589; ss.
OS Ctenocephalides felis.
PN MO9628469-A1.
PD 19-SEP-1996.
PR 08-MAR-1996; U03133.
PR 09-MAR-1995; US-401509.
PA (HESK-) HESKA CORP.
PI Rushlow KE, Stiegler GL;
DR WPI: 96-442861/44.
PT Haematophagous insect calreticulin protein - used to reduce insect
PT infestation and desensitise patients to allergic dermatitis
PS Claim 16: Page 70: 86bp: English.
CC The antisense strand (T39518) of nucleic acid nctCal1589 (T39516)
CC hybridises under stringent conditions to a flea calreticulin gene.
CC nctCal1589 codes for a flea calreticulin protein (W04171) that can
CC be used to alter the blood feeding behaviour of haematophagous
CC insects or to desensitise an animal to allergic dermatitis. The
CC antisense strand, when administered to an animal, reduces
CC calreticulin activity in insects feeding on the animal, thereby
CC reducing the insect burden on the animal and in the environment
CC of the animal.
SQ Sequence 1389 BP: 436 A: 335 C: 266 G: 552 T:

Query Match 57.1%; Score 20; DB 1: Length 1589;
Best Local Similarity 82.1%; Pred. No. 6;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 CAGATCTAGACTAGTCTTCTACACTCT 28
DB 1237 CAGCTCAAGATGCTAGGTTCTACGCTCT 1210

RESULT 5
X20248_08/c
Continuation (9 of 10) of X20248 from base 800001 (Borrelia burgdorferi polynucleotide
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End
WP X20248_00 1 110000
WP X20248_01 210000
WP X20248_02 310000
WP X20248_03 410000
WP X20248_04 510000
WP X20248_05 610000
WP X20248_06 710000
WP X20248_07 810000
WP X20248_08 910000
WP X20248_09 910000
WP X20248_10 910000

Query Match 55.4%; Score 19.4; DB 1: Length 110000;
Best Local Similarity 79.3%; Pred. No. 21;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CAGATCTAGACTAGTCTTCTACACTCT 29
DB 22738 CAGATAAAGATATATTTGTACACTATT 22710

RESULT 6
V59636
ID V59636 standard: DNA: 1291 BP.
AC V59636:
DT 19-JAN-1999 (first entry)
DE Human secreted protein gene 126 clone HELGH31.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

DE Fanconi anaemia of complementation group A gene.
 KW Fanconi anaemia of complementation group A; FA-A; genetic defect;
 KW prenatal FA-A; FA-A carrier detection; disease diagnosis; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 32..4399
 FT /*tag= a
 FT /product= FA-A
 PN WO9814462-A1.
 PD 09-APR-1998.
 PE 03-OCT-1997; U18010.
 PR 04-OCT-1996; US-726012.
 PA (FANC-) FANCONI ANEMIA RES FUND INC.
 PI Joenje H, Lo Ten Foe JR;
 DR MPI; 98-240012/21.
 P-PSDB: M48663.
 PT DNA for Fanconi Anaemia complementation group A - useful for, e.g.
 PT developing products for diagnosis and screening of disease and gene
 PT therapy
 PS Claim 4; Fig 2; 63pp; English.
 CC This sequence encodes the Fanconi anaemia of complementation group A
 CC (FA-A) protein of the invention. The DNA's may be used to complement a
 CC genetic defect in a cell (especially the FA-A gene). The products can be
 CC used for screening (especially prenatal FA-A), detection of FA-A carriers
 CC and FA-A disease diagnosis.
 SO Sequence 5503 BP; 1208 A; 1527 C; 1492 G; 1276 T;

Query Match 53.1%; Score 18.6; DB 1; Length 5503;
 Best Local Similarity 72.7%; Pred. No. 29;
 Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 2 AGATCTAGATCTATGTTCTACACTCTACGTG 34
 Db 5354 AGATCTTAAGCTGCTTTATACACTGCACGTG 5386

RESULT 8
 ID V60751 standard; DNA; 9183 BP.
 AC V60751;
 DT 08-DEC-1998 (first entry)
 DE HIV-1 strain YBF30 complete genomic sequence.
 KW HIV-1 strain YBF30; antibody; oligonucleotide; diagnosis; immunisation;
 OS Human immunodeficiency virus type 1.
 PN FR2756843-A1.
 PD 12-JUN-1998.
 PE 09-DEC-1996; 015087.
 PR 09-DEC-1996; FR-015087.
 PA (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (INSP) INST PASTEUR.
 PI Barre-Sinoussi F, Lousset-Ajaka I, Mauciere P, Saragosti S,
 PI Simon F;
 DR MPI; 98-336114/30.
 PT Non-M, non-O HIV-1 strain YBF30 - useful for diagnosis and
 PT immunisation
 PS Claim 3; Page 19-24; 85pp; French.
 CC This sequence represents the complete genome sequence of the non-M
 CC (major), non-O (outlier) HIV-1 strain YBF30 (CNCM I-1753), isolated from
 CC the Cameroon. The HIV strain, peptides, antibodies and oligonucleotides
 CC derived from it (see V60752-V60798 and W68473-W68482) are used for
 CC diagnosis of or immunisation against non-M, non-O HIV-1 infections.
 CC The oligonucleotides, peptides and antibodies can also be used for
 CC typing HIV strains.
 SO Sequence 9183 BP; 3269 A; 1640 C; 2227 G; 2047 T;

Query Match 53.1%; Score 18.6; DB 1; Length 9183;
 Best Local Similarity 84.0%; Pred. No. 32;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CAGATCTAGATCTATGTTCTACAC 25
 Db 899 CAGATCTAAATAGTATGTTAAATAC 923

RESULT 9
 ID V60753 standard; DNA; 1539 BP.
 AC V60753;
 DT 08-DEC-1998 (first entry)
 DE HIV-1 strain YBF30 gag gene.
 KW HIV-1 strain YBF30; antibody; oligonucleotide; diagnosis; immunisation;
 OS Human immunodeficiency virus type 1.
 FH Key Location/Qualifiers
 FT CDS 1..1539
 FT /*tag= a
 FT /product= "gag protein"
 PN FR2756843-A1.
 PD 12-JUN-1998.
 PE 09-DEC-1996; 015087.
 PR 09-DEC-1996; FR-015087.
 PA (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (INSP) INST PASTEUR.
 PI Barre-Sinoussi F, Lousset-Ajaka I, Mauciere P, Saragosti S,
 PI Simon F;
 DR MPI; 98-336114/30.
 P-PSDB: W68473.
 PT Non-M, non-O HIV-1 strain YBF30 - useful for diagnosis and
 PT immunisation
 PS Claim 3; Page 25-27; 85pp; French.
 CC This sequence represents the gag gene from the non-M (major), non-O
 CC (outlier) HIV-1 strain YBF30 (CNCM I-1753), isolated from the Cameroon.
 CC The HIV strain (see V60751 for complete genome), peptides, antibodies and
 CC oligonucleotides derived from it (see V60752-V60798 and W68473-W68482)
 CC are used for diagnosis of or immunisation against non-M, non-O HIV-1
 CC infections. The oligonucleotides, peptides and antibodies can also be
 CC used for typing HIV strains.
 SO Sequence 1539 BP; 531 A; 314 C; 402 G; 292 T;

Query Match 53.1%; Score 18.6; DB 1; Length 1539;
 Best Local Similarity 84.0%; Pred. No. 24;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CAGATCTAGATCTATGTTCTACAC 25
 Db 551 CAGATCTAAATAGTATGTTAAATAC 575

RESULT 10
 ID 033064 standard; DNA; 1990 BP.
 AC 033064;
 DT 06-MAY-1993 (first entry)
 DE Encodes Babesia bovis 60kD immunoreactive merozoite surface epitope.
 KW babesiosis; cows; cattle; bos taurus; babesia bovis; babesia bigemina;
 OS merozoite; schizont; ss.
 FH Key Location/Qualifiers
 FT cds 1..1990
 FT /*tag= a
 PN US5171685-A.
 PD 15-DEC-1992.
 PE 04-APR-1990; 504461.
 PR 04-APR-1990; US-504461.
 PA (UYFL) UNIV FLORIDA.
 PA (USDA) US SEC OF AGRIC.
 PI Davis WC, Goff WL, Hines SA, Jasmer DP, McElwain TF;
 PI McGuire TC, Palmergh, Perryman LE, Reduker DW;
 DR MPI; 93-008582/01.
 P-PSDB: R30613.

Mon Jan 3 08:07:24 2000

Job time: 8935 sec

us-09-077-173a-3.rng

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 07:20:31 ; Search time 104.56 seconds
(without alignments)
83.748 Million cell updates/sec

Title: US-09-077-173a-4

Perfect score: 35
Sequence: 1 TCTTAAGCTTGAGTACAGTACAGCAAGCTAGT 35

Scoring table: IDENTITY_NUC

Searched: 311585 segs, 125096042 residues

Database : N_Geneseq_36.*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19.6	56.0	237326	1	V57903	Hereditary haemochromatosis
2	19.6	56.0	235033	1	V57926	Hereditary haemochromatosis
3	18	51.4	1927	1	Q47841	Flavonoid-3',5'-hydroxylase
4	18	51.4	2600	1	Q90096	Mouse kappa-3 opsin
5	17.8	50.9	574	1	T18014	Human gene signature
6	17.6	50.3	1537	1	T26895	Bacteriophage phi
7	17.6	50.3	3018	1	V84620	Human secreted protein
8	17.6	50.3	3735	1	V84629	Human secreted protein
9	17.6	50.3	4005	1	V74942	Human secreted protein
10	17.4	49.7	2078	1	T47698	Staphylococcus aureus
11	17.2	49.1	5433	1	Q40945	Banana polyphenol
12	17.2	49.1	133694	1	T13635	Human DNA polymerase
13	17.2	49.1	984	1	T13636	ACNFP genomic DNA
14	17.2	49.1	1679	1	V03349	ACNFP ORF 13, res1
15	17.2	49.1	4854	1	V74417	Home sapiens inter
16	17.2	48.6	17410	1	O53142	Staphylococcus aureus
17	17.2	48.6	17350	1	Q72708	Sequence encoding
18	17.2	48.6	9487	1	Q79140	HOPI human osteog
19	17.2	48.6	17415	1	T11639	Hepatitis C virus
20	17.2	48.6	17410	1	T18381	Human osteogenic p
21	17.2	48.6	6595	1	T94659	HOPI-1 genomic DNA
22	17.2	48.6	17410	1	V15205	Arabidopsis flavon
23	17.2	48.6	9409	1	V52294	Human osteogenic p
24	17.2	48.6	17415	1	V80733	Streptococcus pneu
25	17.2	48.6	17410	1	X00230	Human osteogenic p
26	17.2	48.6	616	1	X21012	Poly nucleotide seq
27	17.2	48.6	134525	1	Q04525	Total base sequenc
28	17.2	48.6	134525	1	Q04525	Total base sequenc
29	17.2	48.6	134525	1	Q04525	Total base sequenc
30	17.2	48.6	134525	1	Q04525	Total base sequenc
31	17.2	48.6	134525	1	Q04525	Total base sequenc
32	17.2	48.6	134525	1	Q04525	Total base sequenc
33	17.2	48.6	134525	1	Q04525	Total base sequenc
34	17.2	48.6	134525	1	Q04525	Total base sequenc
35	17.2	48.6	134525	1	Q04525	Total base sequenc
36	17.2	48.6	134525	1	Q04525	Total base sequenc
37	17.2	48.6	134525	1	Q04525	Total base sequenc
38	17.2	48.6	134525	1	Q04525	Total base sequenc
39	17.2	48.6	134525	1	Q04525	Total base sequenc

C 40	16.6	47.4	793	1	X24101	Plasmid pBBP22 DNA
C 41	16.4	46.9	3744	1	Q10157	Sequence contig. ab
C 42	16.4	46.9	5394	1	T00872	Murine mc26 struct
C 43	16.4	46.9	3269	1	T31291	Rat poly-immunoglob
C 44	16.4	46.9	11000	1	T58840-2	Continuation (3 of
C 45	16.4	46.9	1366	1	T67284	Human urogenital s

ALIGNMENTS

RESULT 1

ID V57903

V57903 standard; DNA; 237326 BP.

AC V57903; 21-DEC-1998 (first entry)

DE Hereditary haemochromatosis subregion from an HH affected individual.

KW Bovine butyrophilin; BT: human hereditary haemochromatosis; HFE:

KW diagnosis; iron metabolism; NPT3; NPT4; Roret; BRF1; BRF2; BRF3;

KW BRF4; BRF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;

KW type 1 sodium transport gene; ss.

OS Homo sapiens.

PN WO9814466-A1.

PD 09-APR-1998.

PF 30-SEP-1997; U17658.

PR 07-MAY-1997; US-852495.

PR 01-OCT-1996; US-724394.

PA (PROG-) PROGENITOR INC.

PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,

PI Tsuchihashi Z, Wolff RK;

DR WPI: 98-240014/21.

PT Hereditary haemochromatosis gene products - used to develop products

PT for the diagnosis and treatment of hereditary disorders in iron

PT metabolism

PS Claim 1: Fig 9: 209bp; English.

CC The present invention describes hereditary haemochromatosis gene

CC products from the human haemochromatosis gene. The present sequence

CC represents a hereditary haemochromatosis subregion from an hereditary

CC haemochromatosis (HH) affected individual. Also described is a

CC method to determine the presence or absence of the common hereditary

CC haemochromatosis (HFE) gene mutation in an individual comprising:

CC (a) providing DNA or RNA from the individual; and (b) assessing the

CC DNA or RNA for the presence or absence of a haplotype or genotype where

CC the presence or absence of the haplotype genotype indicates the likely

CC presence of the HFE gene mutation in the genome of the individual. The

CC HFE gene sequences from the present invention can be used to develop

CC products for use in the diagnosis and treatment of HFE. The present

CC invention also describes BRF genes, which are homologues of the milk

CC protein butyrophilin (BRF), and can be used in the production of agonists

CC and antagonists of BRF function. Also described are: (1) a Roret gene

CC which can be used to develop products for the study, diagnosis and

CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes

CC which are homologues of a type 1 sodium transport gene, and can

CC similarly be used for hypophosphatemia.

CC Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T;

SQ

Query Match 56.0%; Score 19.6; DB 1; Length 237326;

Best Local Similarity 73.5%; Pred No. 10;

Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 CTTAAGCTTGAGTACAGTACAGCAAGCTAGT 35

DB 83534 CTTAAGCTTGAGTACAGTACAGCAAGCTAGT 83567

RESULT 2

ID V57926

V57926 standard; DNA; 235033 BP.

AC V57926; 23-DEC-1998 (first entry)

DE Hereditary haemochromatosis subregion from an unaffected individual.

KW Bovine butyrophilin; BT: human hereditary haemochromatosis; HFE:

23	20	57.1	415	47	A1506876
24	20	57.1	874	80	A0560640
25	20	57.1	517	89	A0757927
26	19.8	56.6	435	29	AA143686
27	19.8	56.6	471	42	AA115953
28	19.8	56.6	432	42	AA115954
29	19.8	56.6	560	79	A0313720
30	19.6	56.0	405	29	AA162990
31	19.6	56.0	424	48	AA160395
32	19.6	56.0	1101	69	CNS00022
33	19.6	56.0	446	78	A0404299
34	19.6	56.0	492	90	A0803121
35	19.4	55.4	448	30	A0803131
36	19.4	55.4	200	40	AA636172
37	19.4	55.4	499	72	AI647109
38	19.4	55.4	522	81	A061367
39	19.4	55.4	574	90	A0786710
40	19.2	54.9	226	20	T18141
41	19.2	54.9	377	20	T18637
42	19.2	54.9	432	26	W39972
43	19.2	54.9	445	28	AA119812
44	19.2	54.9	433	28	AA120292
45	19.2	54.9	579	34	AA511097
					A1506876
					V156601.X
					A0560640
					HS_2066.B
					A0757927
					HS_5506.B
					AA143686
					z607C05
					AA115953
					u69605.X
					Y1155954
					us96606.Y
					Y1155954
					us96606.Y
					A0313720
					RFC111-95
					AA162990
					ms16605.X
					AI603956
					ms16605.Y
					AI050452
					Drosophil
					A0404299
					HS_5069.B
					A0803131
					HS_3158.B
					AA636172
					Vn15603.X
					AI647109
					Vn15603.Y
					AI080824
					HS_3220.B
					A061367
					HS_5088.B
					A0786710
					HS_3189.A
					T18141
					0707C3
					czap
					T18637
					5c07a05-l7
					W39972
					412 Mouse
					V
					AA119812
					mm14c08.X
					AA120292
					mm16c08.X
					AA511097
					Vn61502.X

ALIGNMENTS

BASE COUNT	154 a	84 c	71 g	153 t
ORIGIN	RPC111 Human Male BAC library"			
Query Match	62.9% Score 22; DB 80; Length 462;			
Best Local Similarity	83.3% Pred. No. 25;			
Matches 25; Conservative	0; Mismatches 5; Indels 0; Gaps 0;			
QY	1	CAGACTGATAGTACTATGTTCTACACTCTTA	30	
Db	183	CAGACTCGATGCTATGTTCTACACTCTTA	212	
RESULT	2			
LOCUS	T79250			
DEFINITION	T79250	473 bp	mRNA	EST
	y470c09.r1 Soares fetal liver spleen INFLS Homo sapiens CDNA clone			
ACCESSION	T79250			
	IMAGE:113584 5', mRNA sequence.			
NID	G687759			
VERSION	T79250.1	GI:697759		
KEYWORDS	EST.			

```

RESULT      1
LOCUS       A0530999
DEFINITION  A0530999 462 bp DNA GSS 17-MAY-1999
ACCESSION  RPCI-11-379C24.TV RPCI-11 Homo sapiens genomic clone
NID        RPCI-11-379C24, genomic survey sequence.
KEYWORDS   A0530999
SOURCE     A0530999.1 GI:4843042
ORGANISM   human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 462)
AUTHORS    Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
TITLE       Use of BAC End Sequences from library RPCI-11 for Sequence-Ready
            Map Building
JOURNAL     Unpublished (1997)
COMMENT     Other-GSSs: RPCI-11-379C24.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbeetlgr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieterdejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering/) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: 27
            Class: BAC ends.

FEATURES             source
     location/Qualifiers
         1..462
         /organism="Homo sapiens"
         /db_xref="GDB:7645223"
         /db_xref="taxon:9606"
         /clone="RPCI-11-379C24"
         /clone_1lb="RPCI-11"
         /sex="Male"
         /cell_type="lymphocytes"
         /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;"

```

```

TITLE
JOURNAL
COMMENT
The Mashu-Merck EST Project
Unpublished (1995)
Other_ESTS: yd70c09.s1
Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 688
High quality sequence stops: 363
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 688 Std Error: 0.00
Seq primer: M13RPI
High quality sequence stop: 363.
Location/Qualifiers
1. .473
/organism="Homo sapiens"
/db_xref="GDB:469201"
/db_xref="taxon:9606"
/clone="IMAGE:113584"
/clone_1id="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
15' ACTGCAGCAATTAATTAAGATCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bonaldo."
BASE COUNT
135 a 73 c 76 g 185 t 4 others
Query Match
59.4%; Score 20.8; DB 21; Length 473;

```

Query Match

59.48; Score 20.8; DB 21; Length 473;

Best Local Similarity 78.1%; Pred. No. 75;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 AGACTAGATCTATGTTCTACACTCTTACGT 33
Db 313 ACATCTACCTACTATGATCTACAAATTTAGGT 344

RESULT 3
AA258261 411 bp mRNA EST 06-AUG-1997
LOCUS z595910.r1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:667746
DEFINITION 5', mRNA sequence.
ACCESSION AA258261
NID g1893423
VERSION AA258261.1 GI:1893423
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 411)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, K., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393385.

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 708 Std Error: 0.00
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 397.

FEATURES

source
1..411
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:667746"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NDH, pregnant uterus NBHPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 26032-265223, 340488-345479, and 484488-489479."
BASE COUNT 118 a 61 c 62 g 170 t
ORIGIN

Query Match 59.4%; Score 20.8; DB 30; Length 411;
Best Local Similarity 78.1%; Pred. No. 76;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 AGACTAGATCTATGTTCTACACTCTTACGT 33
Db 256 ACATCTACCTACTATGATCTACAAATTTAGGT 287

RESULT 4

AA465759 419 bp mRNA EST 11-JUN-1997
LOCUS v107e06.r1 Knowles Solter mouse blastocyst B3 Mus musculus cDNA
DEFINITION clone IMAGE:835042 5', mRNA sequence.
ACCESSION AA465759
NID g2191899
VERSION AA465759.1 GI:2191899
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 419)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Nov 29, 1993 this sequence version replaced gi:503248.

TITLE
JOURNAL
COMMENT

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouse@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:495258
High quality sequence stop: 396.

FEATURES

source
1..419
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/map="13q12"
/clone="IMAGE:835042"
/clone_lib="Knowles Solter mouse blastocyst B3"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/note="Organ: embryo; Vector: pSPORT; Site_1: NotI; Site_2: SalI; Cloned unidirectionally from mRNA prepared from 800 blastocysts. Primer: SalI(5'):
5'-CGGTCGACCGTCGACCGTTTCTTTTCTTT-3'. cDNAs were cloned into the NotI/SalI sites of a pSPORT vector (Life Technologies) and B3."
BASE COUNT 120 a 74 c 83 g 142 t
ORIGIN

Query Match 59.4%; Score 20.8; DB 34; Length 419;
Best Local Similarity 78.1%; Pred. No. 76;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 AGACTAGATCTATGTTCTACACTCTTACGT 33
Db 61 AGATCTAGATGTTGCTTCTTACACTCTTACGT 92

RESULT 5
AA669879/c 442 bp mRNA EST 12-MAY-1999
DEFINITION ag42d01.s1 Jia Done marrow stroma Homo sapiens cDNA clone
IMAGE:1119457 3', mRNA sequence.
ACCESSION AA669879
NID g2631378

```

VERSION      AA669879.1  GI:2631378
KEYWORDS     EST
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE        1 (bases 1 to 442)
JOURNAL      Jia,L., Robery,P., Young,M., Schuler,G., Powell,J., Yang,L.,
COMMENT      Lennon,G., Hillier,L., Allen,M., Bowles,L., Gelsel,S., Kucaba,T.,
              Marra,M., Martin,J., Steptoe,M., Tan,F., Theising,B., Bowers,Y.,
              Wylie,T., Waterston,R., Wilson,R. and Francomano,C.
              WashU-MGB/NHGRI EST Project
              Unpublished (1997)
              On Nov 6, 1997 this sequence version replaced gi:933403.

Contact: Wilson RK / Jia L
WashU-MGB/NHGRI EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40m3 fwd. Et from Amersham
High quality sequence stop: 400.
Location/Qualifiers
1..442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1119457"
/clone_11b="Jia Bone marrow stroma"
/sex="mixed"
/tissue_type="Bone marrow stroma"
/dev_stage="mixed"
/lab_host="XLI-Blue MRF/SOLR"
/note="vector: pBluescript; Site_1: EcoRI; Site_2: XhoI;
mRNA made from human bone marrow stroma, cDNA made by
O190-dT priming. Directionally cloned. Size-selected for
average insert size >0.5 kb. Library supplied by Dr. Libin
Jia (NHGRI)."
```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdip/image/image.html

Insert length: 902 Std Error: 0.00
Seq primer: -40m13 fwd, ET from Amersham
High quality sequence stop: 279.
Location/Qualifiers
1. 326
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1350470"
/clone_1ld="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I, Site_2: Eco RI, 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTACCAATGCAAGTGGAGGCGCCGCTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN
107 a 43 c 49 g 127 t

Query Match 59.4%; Score 20.8; DB 38; Length 326;
Best Local Similarity 78.1%; Pred. No. 79;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 AGATCTAGTACTATGTTCTCACTCTTACGT 33
||||| ||||||| ||||||| ||| ||
Db 321 ACATCTACTACTATGATCTCAATTTTACGT 290

RESULT 7
A1378554/c
LOCUS
DEFINITION
A1378554 485 bp mRNA EST 18-MAR-1999
tc72a02.x1 Soares_nhmhu_SI Homo sapiens cDNA clone IMAGE:2070122
3', mRNA sequence.
ACCESSION
A1378554
NID
94188407
VERSION
A1378554.1 GI:4188407
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 485)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Oct 8, 1998 this sequence version replaced gi:3726019.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 449 Std Error: 0.00
 Seq primer: -40UP from Glibco
 High quality sequence stop: 443.
 Location/Qualifiers

FEATURES

1. 485
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2070122"
 /clone_lib="Soares_NHMPU.S1"
 /tissue_type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: pRT73D-Pac
 (Pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NBHM, pregnant uterus
 NBHPU, and fetal heart NBH19W) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of 1.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."
 194 a 75 c 72 g 144 t

BASE COUNT

ORIGIN

Query Match 59.4%; Score 20.8; DB 45; Length 485;
 Best Local Similarity 78.1%; Pred. No. 74;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 AGATCTAGTACTATGTTCTACACTTACGT 33
 Db 163 ACATCTACCTACTATGATCTACAAATTAGCT 132

RESULT 8
 A1754738 527 bp mRNA EST 22-JUN-1999
 LOCUS cr29604.x1 jia bone marrow stroma Homo sapiens cDNA clone
 DEFINITION HMSC.cr29604 3', mRNA sequence.
 ACCESSION A1754738
 NID 95133002
 VERSION A1754738.1 GI:5133002
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 527)
 Jia, L.B., Young, M.F., Touchman, J.W., Bouffard, G.G.,
 Beckstrom-Sternberg, S.M., Green, E.D., Powell, J.T., Yang, L.M.,
 Robey, P.G., Hotchkiss, R.N. and Francmanno, C.A.
 SGA: The Skeletal Genome Anatomy Project
 Unpublished (1997)
 On Dec 20, 1995 this sequence version replaced gi:1135954.

TITLE

JOURNAL

COMMENT

CONTACT: Libin Jia
 Medical Genetics Branch
 National Human Genome Research Institute
 10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
 Tel: 301-402-4877
 Fax: 301-496-7157
 Email: libin@helix.nih.gov
 DNA Sequencing and analyses by National Institutes of Health
 Intramural Sequencing Center (NISC).
 Plate: 29 row: e column: 04
 Seq primer: -21M13 forward primer (ABI).
 Location/Qualifiers
 1. 527

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HMSC.cr29604"
 /clone_lib="jia bone marrow stroma"
 /sex="mixed"
 /tissue_type="bone marrow stroma"
 /dev_stage="mixed"
 /lab_host="XLI-Blue MRF/SOLR"
 /note="Vector: pBluescript; Site_1: EcoRI; Site_2: XhoI;
 mRNA made from human bone marrow stroma, cDNA made by
 oligo-dT priming. Directionally cloned. Size-selected for
 average insert size >0.5 kb. Library supplied by Dr. Libin
 Jia (NHGRI)."
 195 a 74 c 77 g 180 t 1 others

BASE COUNT

ORIGIN

Query Match 59.4%; Score 20.8; DB 51; Length 527;
 Best Local Similarity 78.1%; Pred. No. 73;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 AGATCTAGTACTATGTTCTACACTTACGT 33
 Db 200 ACATCTACCTACTATGATCTACAAATTAGCT 169

RESULT 9
 A1769109 318 bp mRNA EST 28-JUN-1999
 LOCUS wg32h02.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:2366835 3', mRNA sequence.
 ACCESSION A1769109
 NID 95235618
 VERSION A1769109.1 GI:5235618
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 318)
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 Tumor Gene Index
 Unpublished (1997)
 On Dec 20, 1995 this sequence version replaced gi:1135517.

TITLE

JOURNAL

COMMENT

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Glibco
 High quality sequence stop: 272.
 Location/Qualifiers
 1. 318
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="17c"
 /clone="IMAGE:2366835"
 /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pRT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and ss circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and clones: Soares NBHSF pool 1:
 309384-310919, 323208-325895 Soares NB2HP pool 1:
 145032-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NB2HF-9W pool 1:

NID 95884786
 VERSION A1981758.1 GI:5884786
 KEYWORDS EST.
 SOURCE chicken.
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
 Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 1 (bases 1 to 623)
 REFERENCE Titunagaru,V.G., Sofer,L. and Burnside,J.
 An expressed-sequence-tag database of activated chicken T cells:
 sequence analysis of 5596 clones
 Unpublished (1999)
 JOURNAL On May 18, 1998 this sequence version replaced gi:3136799.
 COMMENT

FEATURES
 source
 1..623
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /map="4; 4; 21q"
 /clone="pat.pk0064.e2.f"
 /clone.lib="chicken activated T cell cDNA"
 /sex="male"
 /cell_type="Con A-activated splenic T cell"
 /lab_host="E.coli TOP10 F"
 /note="Vector: pCDNA3"
 Seq primer: 77.
 Location/Qualifiers

BASE COUNT 218 a 101 c 145 g 153 t 6 others
 ORIGIN

Query Match 58.9%; Score 20.6; DB 63; Length 623;
 Best Local Similarity 74.3%; Pred. No. 86;
 Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CAGATCTAGTACTATGCTCTACTCTACGTCG 35
 Db 523 CAGATCCAGATGATCTCTCTCTACGTCG 489

RESULT 13
 A0290000 582 bp DNA GSS 02-DEC-1998
 LOCUS nbxb0036M03f CUG1 Rice BAC Library Oryza sativa genomic clone
 DEFINITION nbxb0036M03f, genomic survey sequence.
 ACCESSION A0290000
 NID 93951362
 VERSION A0290000.1 GI:3951362
 KEYWORDS GSS.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Charophyta; Embryophyta group;
 Magnoliophyta; Tracheophyta; euphyllophytes; Spermatophyta;
 Magnoliopsida; Liliopsida; Poales; Poaceae; Oryza.
 1 (bases 1 to 582)
 REFERENCE Yu,Y., Budiman,M.A., Boyd,C.A., Gaudette,B., Gayle,A.,
 Kingsbury,R., Phillips,K., Sasinowski,M., Dean,R.A. and Wing,R.A.
 A BAC End Sequencing Framework to Sequence the Rice Genome
 Unpublished (1998)
 JOURNAL
 TITLE
 COMMENT

Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu

Seq primer: TAATACGACTCACTATAGG
 Class: BAC ends
 High quality sequence stop: 483.
 Location/Qualifiers

FEATURES
 source
 1..582
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="nbxb0036M03f"
 /clone.lib="CUG1 Rice BAC Library"
 /tissue="leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBelBAC11; Site_1: HindIII; Site_2:
 HindIII; Rice is one of two most popular grains in the
 world. Half of the world population especially those
 inhabiting highly populated areas of the humid tropics
 and subtropics, rely on rice as their primary source of
 carbohydrate. Monocotyledonous rice is a diploid plant
 (2n=24) with a haploid genome equivalent of 431 Mb
 (Arumuganathan and Earle, 1991). The relatively small
 genome of rice, three times larger than that of
 Arabidopsis, makes it suitable for genomic studies. In
 order to facilitate positional cloning, physical mapping
 and genome sequencing of rice, we have constructed a BAC
 library from Oryza sativa, Nipponbare variety. The
 library contains 36,864 clones with an average insert size
 of 128.5 Kb providing 10.9 haploid genome equivalents.
 The deep coverage allows the isolation a particular
 sequence with a probability of 99.9%. Two high density
 filters, each containing 18,432 clones (doubly spotted),
 represent the whole library for colony screening."

BASE COUNT 169 a 95 c 131 g 187 t
 ORIGIN

Query Match 58.9%; Score 20.6; DB 77; Length 582;
 Best Local Similarity 85.2%; Pred. No. 87;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 ATCTAGTACTATGCTCTACTCTTA 30
 Db 395 ATGTAGTACTATGCTCTATATCTTA 369

RESULT 14
 A0509564 545 bp DNA GSS 03-MAY-1999
 LOCUS nbxb0096H07r CUG1 Rice BAC Library Oryza sativa genomic clone
 DEFINITION nbxb0096H07r, genomic survey sequence.
 ACCESSION A0509564
 NID 94733647
 VERSION A0509564.1 GI:4733647
 KEYWORDS GSS.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Oryza.
 1 (bases 1 to 545)
 REFERENCE Yu,Y., Budiman,M.A., Boyd,C.A., Gaudette,B., Gayle,A.,
 Kingsbury,R., Phillips,K., Sasinowski,M., Dean,R.A. and Wing,R.A.
 A BAC End Sequencing Framework to Sequence the Rice Genome
 Unpublished (1998)
 JOURNAL
 TITLE
 COMMENT

Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: GGAACACCTATGACCATG

```

/*tag= b
/notes= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

```

```

EP-786519-A2.
30-JUL-1997.
07-JAN-1997; 100117.
05-JAN-1996; US-009861.
(HUMA-) HUMAN GENOME SCI INC.
Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
Rosen CA;
WPI; 97-374922/35.
Polynucleotide(s) and proteins derived from Staphylococcus aureus
stored on computer readable medium and used in the production of
anti-S.aureus vaccines
PS Claim 1; Page 620-622; 3271pp; English.
This sequence represents one of 5191 Staphylococcus aureus DNA sequences
of the invention. The DNA sequences are recorded on a computer readable
medium, preferably selected from a floppy or hard disk, random access
memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
the S.aureus DNA sequences allows putative functions to be assigned so
that protein-encoding or regulatory regions of commercial, therapeutic or
industrial importance can be obtained. Specifically, sequences which are
likely to encode antigens have been identified and these polypeptides can
be used in a vaccine composition against S.aureus infection. The
polypeptides can also be used in a kit for the immunodetection of
S.aureus in a sample. S.aureus is implicated in numerous human diseases,
including cellulitis, eyelid infections, food poisoning, osteomyelitis,
skin and surgical wound infections, scalded skin syndrome, toxic shock
syndrome, etc. Organisms transformed with the DNA sequences can be used
for recombinant production of the polypeptides. The new DNA sequences
(and their fragments) are useful as primers or probes for isolating
homologues of any of the S.aureus DNA sequences contained on the
computer readable medium.
SO Sequence 4854 BP; 1615 A; 650 C; 986 G; 1480 T;

```

```

Query Match 49.1%; Score 17.2; DB 1; Length 4854;
Best Local Similarity 71.0%; Pred. No. 67;
Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

OY 4 TAAGCTTGAGTCAGCAGCAGCACTACT 34
| | | | | | | | | | | | | | | |
DB 2908 TGAGCTTGAGCTTATCTAGCAGACGATT 2938

```

Search completed: December 29, 1999, 07:20:52
Job time: 8956 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 04:26:25 ; Search time 727.51 seconds
(without alignments)
164.915 Million cell updates/sec

Title: US-09-077-173a-3
Perfect score: 35
Sequence: 1 CAGATCTAGTACTATGTTCTACACTCTTACGTC 35

Scoring table: IDENTITY_NUC

Searched: 4089388 seqs, 1713965092 residues

Database: EST:*

Word size: 0

Number of hits that pass the threshold: 8178776

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gp_est1:*
- 21: gp_est2:*
- 22: gp_est3:*
- 23: gp_est4:*
- 24: gp_est5:*
- 25: gp_est6:*
- 26: gp_est7:*
- 27: gp_est8:*
- 28: gp_est9:*
- 29: gp_est10:*
- 30: gp_est11:*
- 31: gp_est12:*
- 32: gp_est13:*
- 33: gp_est14:*
- 34: gp_est15:*
- 35: gp_est16:*
- 36: gp_est17:*
- 37: gp_est18:*
- 38: gp_est19:*
- 39: gp_est20:*
- 40: gp_est21:*
- 41: gp_est22:*
- 42: gp_est23:*
- 43: gp_est24:*
- 44: gp_est25:*
- 45: gp_est26:*
- 46: gp_est27:*
- 47: gp_est28:*
- 48: gp_est29:*
- 49: gp_est30:*

- 50: gp_est31:*
- 51: gp_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gp_est33:*
- 60: gp_est34:*
- 61: gp_est35:*
- 62: gp_est36:*
- 63: gp_est37:*
- 64: gp_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gp_gss1:*
- 70: gp_gss2:*
- 71: gp_gss3:*
- 72: gp_gss4:*
- 73: em_gss1:*
- 74: em_gss2:*
- 75: em_gss3:*
- 76: em_gss4:*
- 77: gp_gss5:*
- 78: gp_gss6:*
- 79: gp_gss7:*
- 80: gp_gss8:*
- 81: gp_gss9:*
- 82: em_gss5:*
- 83: em_gss6:*
- 84: em_gss7:*
- 85: em_gss8:*
- 86: em_gss9:*
- 87: em_gss10:*
- 88: em_gss11:*
- 89: gp_gss10:*
- 90: gp_gss11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	22	62.9	462	80	A0530999	A0530999 RPT-11-3
2	20.8	59.4	473	21	T79250	T79250 yd70c09.r1
3	20.8	59.4	411	30	AA258261	AA258261 z15910.r
4	20.8	59.4	419	34	AA465759	AA465759 v107e06.r
5	20.8	59.4	442	37	AA669879	AA669879 ag42d01.s
6	20.8	59.4	326	38	AA806372	AA806372 cc22e08.s
7	20.8	59.4	485	45	AI378554	AI378554 tc72d02.x
8	20.8	59.4	527	51	AI754738	AI754738 cr29e04.x
9	20.8	59.4	489	59	AI769109	AI769109 wg32h02.x
10	20.8	59.4	489	70	A0002323	A0002323 CIT-HSP-2
11	20.6	58.9	1311	25	W15735	W15735 mb53903.r1
12	20.6	58.9	623	63	AI981758	AI981758 pat.pk006
13	20.6	58.9	582	77	A0290000	A0290000 nbx0036M
14	20.6	58.9	543	79	A0309564	A0309564 nbx00096H
15	20.6	58.9	364	81	A0690006	A0690006 nbx00081G
16	20.6	58.9	207	90	A0794814	A0794814 nbx00054G
17	20.4	58.3	431	71	A0111301	A0111301 CIT-HSP-2
18	20.2	57.7	439	22	R49688	R49688 y664d11.s1
19	20.2	57.7	396	70	A0010184	A0010184 HS-2263.B
20	20.2	57.7	371	71	A0104808	A0104808 HS-2166.B
21	20.2	57.7	475	77	A0296901	A0296901 HS-3005.B
22	20.2	57.7	466	79	A0493046	A0493046 HS-5222.B

KM diagnosis; iron metabolism; NPT3; NPT4; RORet; BTF1; BTF2; BTF3;
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
 KM type 1 sodium transport gene; ss.
 OS Homo sapiens.
 PN WO9814466-A1.
 PD 09-APR-1998.
 PF 30-SEP-1997; U17658.
 PR 07-MAY-1997; US-852495.
 PR 01-OCT-1996; US-724394.
 PA (PROG-) PROGNITOR INC.
 PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WT,
 PI Tsuchihashi Z, Wolf RK;
 DR WPI: 98-240014/21.
 PT Hereditary haemochromatosis gene products - used to develop products
 PT for the diagnosis and treatment of hereditary disorders in iron
 PT metabolism
 PS Example 2; Fig 8; 209pp; English.
 CC The present invention describes hereditary haemochromatosis gene
 CC products from the human haemochromatosis gene. The present sequence
 CC represents a hereditary haemochromatosis subregion from an individual
 CC unaffected by hereditary haemochromatosis (HH). Also described is a
 CC method to determine the presence or absence of the common hereditary
 CC haemochromatosis (HFE) gene mutation in an individual comprising:
 CC (a) providing DNA or RNA from the individual; and (b) assessing the
 CC DNA or RNA for the presence or absence of a haplotype or genotype where
 CC the presence or absence of the haplotype genotype indicates the likely
 CC presence of the HFE gene mutation in the genome of the individual. The
 CC HFE gene sequences from the present invention can be used to develop
 CC products for use in the diagnosis and treatment of HFE. The present
 CC invention also describes BTF genes, which are homologues of the milk
 CC protein butyrophilin (BT), and can be used in the production of agonists
 CC and antagonists of BT function. Also described are: (1) a RORet gene
 CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia.
 SQ Sequence 235033 BP; 68786 A; 48466 C; 49441 G; 68340 T;

Query Match 56.0%; Score 19.6; DB 1; Length 235033;
 Best Local Similarity 73.5%; Pred. No. 10;
 Matches 23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 CTTAGCTTGAGTCAGTACGACGACGACTGTT 35
 ||||||| ||||| |||||||
 Db 83581 CTAAGCTTGCTGCTCATTAAGAGTGTACTGACTT 83614

RESULT 3
 ID 047841 standard; mRNA to cDNA; 1927 BP.
 AC 047841;
 DT 28-MAR-1994 (first entry)
 DE Flavonoid-3',5'-hydroxylase gene.
 KW Flavonoid-3',5'-hydroxylase; transformation; plants; petunia; rose;
 KM tobacco; pigment alteration; blue; SSP; single specific primer; ss.
 OS Campanula medium.
 FH Key Location/Qualifiers
 FT cds 180..1751
 FT /*tag- a
 FT /product= Flavonoid-3',5'-hydroxylase
 PN WO9318155-A.
 PD 16-SEP-1993.
 PF 20-NOV-1992; J01520.
 PR (XOM) KYOWA HAKKO KOGYO CO LTD.
 PR (YOM) KYOWA HAKKO KOGYO CO LTD.
 PI Kikuchi Y, Kiyokawa S, Ohbayashi M, Okinaka Y, Shimada R;
 PI Shimada Y;
 DR WPI: 93-303469/38.
 DR P-PSDB; R40868.
 PT Gene coding for flavonoid-3',5'-hydroxylase of petunia petals -
 PT used to transform plants e.g. petunia, rose or tobacco to give
 PT bluer flower colour and altered pigment pattern

PS Claim 1; Page 71-75; 82pp; Japanese.
 CC Insertion of the sequences (047840-42) into plants such as rose,
 CC petunia, tobacco and carnation, using a suitable vector such as
 CC agrobacterium, give transformed plants which express the gene,
 CC resulting in petals with a bluer colour than normal, and/or
 CC pigmentation patterns which do not occur naturally. The sequences
 CC were amplified using primers (047843-70). Related single specific
 CC primers using a gene sequence coding for the haem-binding region of
 CC cytochrome P450 are shown in (047871-047903).
 SQ Sequence 1927 BP; 582 A; 399 C; 396 G; 550 T;

Query Match 51.4%; Score 18; DB 1; Length 1927;
 Best Local Similarity 70.6%; Pred. No. 25;
 Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 TCTTAGCTTGAGTCAGTACGACGACGACTGTT 34
 ||||| || ||||||| |||||
 Db 656 TCATATGCTCAAAAGAAATGTACGACGACATCGAGT 689

RESULT 4
 ID 090096 standard; cDNA; 2600 BP.
 AC 090096;
 DT 03-NOV-1995 (first entry)
 DE Mouse kappa-3 oploid receptor.
 KM Kappa-3 oploid receptor; analgesia; ss.
 OS Mus sp.
 FH Key Location/Qualifiers
 FT cds 299..1402
 FT /*tag- a
 FT WO9512616-A.
 PN PD 11-MAY-1995.
 PF 03-NOV-1994; U12728.
 PR 05-NOV-1993; US-147949.
 PA (SILOK) SLOAN KETTERING INST CANCER RES.
 PI Pan Y, Pasternak GW;
 DR WPI: 95-193814/25.
 DR P-PSDB; R74298.
 DT Nucleic acid molecule(s) encoding a kappa-3 oploid receptor, and
 PT antibody against the receptor - used to detect the receptor, and to
 PT image cell membrane-bound receptor in a subject
 PS Disclosure; Fig.1; 68pp; English.
 CC Degenerate primers based on conserved sequences of the mouse delta
 CC oploid receptor were used in PCR to amplify mouse kappa-3 oploid
 CC receptor cDNA from a mouse brain lambda ZAP cDNA library.
 SQ Sequence 2600 BP; 536 A; 680 C; 663 G; 721 T;

Query Match 51.4%; Score 18; DB 1; Length 2600;
 Best Local Similarity 80.8%; Pred. No. 26;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTTAGCTTGAGTCAGTACGACGACGACTGTT 27
 ||||||| ||||| |||||||
 Db 1778 CTTAGCTTGAGTCAGTACGACGACGACTGTT 1803

RESULT 5
 ID T19014 standard; cDNA to mRNA; 574 BP.
 AC T19014;
 DT 14-JUN-1996 (first entry)
 DE Human gene signature HUMS00015.
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KM human; cloning; mapping; non-biased library; diagnosis; detection;
 KM cell typing; abnormal cell function; ss.
 OS Homo sapiens.
 PN WO9514772-A1.
 PD 01-JUN-1995.
 PF 11-NOV-1994; J01916.
 PR 12-NOV-1993; JP-355504.

CC 06-JUN-1997; US-048962.
PR 06-JUN-1997; US-048970.
PR 06-JUN-1997; US-048974.
PR 06-JUN-1997; US-049373.
PR 05-SEP-1997; US-057584.
PR 05-SEP-1997; US-057629.
PR 05-SEP-1997; US-057642.
PR 05-SEP-1997; US-057645.
PR 05-SEP-1997; US-057648.
PR 05-SEP-1997; US-057651.
PR 05-SEP-1997; US-057766.
PR 06-JUN-1997; US-048876.
PR 06-JUN-1997; US-048880.
PR 06-JUN-1997; US-048883.
PR 06-JUN-1997; US-048892.
PR 06-JUN-1997; US-048895.
PR 06-JUN-1997; US-048898.
PR 06-JUN-1997; US-048901.
PR 06-JUN-1997; US-048917.
PR 06-JUN-1997; US-048963.
PR 06-JUN-1997; US-048971.
PR 06-JUN-1997; US-049019.
PR 06-JUN-1997; US-049374.
PR 05-SEP-1997; US-057627.
PR 05-SEP-1997; US-057634.
PR 05-SEP-1997; US-057643.
PR 05-SEP-1997; US-057646.
PR 05-SEP-1997; US-057649.
PR 05-SEP-1997; US-057654.
PR 05-SEP-1997; US-057666.
PR 05-SEP-1997; US-057760.
PR 05-SEP-1997; US-057763.
PR 05-SEP-1997; US-057769.
PR 05-SEP-1997; US-057774.
PR 05-SEP-1997; US-057777.
PR (HUMA-) HUMAN GENOME SCI. INC.
PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,
PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
DR WPI: 99-059865/05.
DR P-PSDB: W88743.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 4: Page 472-473; 772pp; English.
CC The invention relates to nucleic acid sequences (Y84411 to Y84633)
CC encoding human secreted proteins (W88534 to W88756). The secreted protein
CC gene sequences are deposited with the ATCC under deposit numbers ATCC
CC 97979, 97976, 97975, 97977, 209007, 209008, 209009, 209010.
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC cells comprising recombinant vectors containing the nucleic acid
CC sequences are used for the recombinant production of the secreted
CC proteins. The polynucleotide and amino acid sequences are useful for are
CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein or gene therapy. Pathological conditions can be also diagnosed
CC by determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, neurodegenerative
CC disorders, developmental abnormalities and foetal deficiencies, blood
CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
CC such as osteoporosis, arthritis or malignancies, diseases of testes, lungs

CC	or thymus; digestive/endocrine disorders; infections and AIDS. The
CC	polypeptides are also useful for identifying their binding partners.
CC	The present sequence represents a gene encoding a human secreted protein
CC	(see descriptor line for gene number and clone identification).
SC	Sequence 3018 BP; 841 A; 635 C; 678 G; 863 T;
OY	3 TTAACCTGGAGTCACGTACGACGAACTACT 34
ID	1925 TAAATCTGGAGCACCAGTACGTGCMAAAGTG 1894
RESULT	8
VB44429/c	
ID	VB44429 standard; DNA; 3735 BP.
AC	V84429:
DT	01-MAR-1999 (first entry)
DE	Human secreted protein gene 19 clone HSAVU34.
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW	inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS	Homo sapiens.
PN	WO9854963-A2.
PD	10-DEC-1998.
PE	04-JUN-1998; U11422.
PR	18-DEC-1997; US-070923.
PR	06-JUN-1997; US-048877.
PR	06-JUN-1997; US-048881.
PR	06-JUN-1997; US-048883.
PR	06-JUN-1997; US-048893.
PR	06-JUN-1997; US-048896.
PR	06-JUN-1997; US-048899.
PR	06-JUN-1997; US-048915.
PR	06-JUN-1997; US-048949.
PR	06-JUN-1997; US-048964.
PR	06-JUN-1997; US-048972.
PR	06-JUN-1997; US-049020.
PR	06-JUN-1997; US-049375.
PR	05-SEP-1997; US-057628.
PR	05-SEP-1997; US-057635.
PR	05-SEP-1997; US-057644.
PR	05-SEP-1997; US-057647.
PR	05-SEP-1997; US-057650.
PR	05-SEP-1997; US-057661.
PR	05-SEP-1997; US-057761.
PR	05-SEP-1997; US-057764.
PR	05-SEP-1997; US-057770.
PR	05-SEP-1997; US-057775.
PR	05-SEP-1997; US-057778.
PR	06-JUN-1997; US-048875.
PR	06-JUN-1997; US-048878.
PR	06-JUN-1997; US-048882.
PR	06-JUN-1997; US-048885.
PR	06-JUN-1997; US-048894.
PR	06-JUN-1997; US-048897.
PR	06-JUN-1997; US-048900.
PR	06-JUN-1997; US-048916.
PR	06-JUN-1997; US-048962.
PR	06-JUN-1997; US-048970.
PR	06-JUN-1997; US-048974.
PR	06-JUN-1997; US-049373.
PR	05-SEP-1997; US-057584.
PR	05-SEP-1997; US-057629.

DB 3504 TAAGTTGCAATCACTATGACCACTCTCTT 3535

RESULT 10

TT47698

TT47698 standard; DNA: 2078 BP.

DE 12-AUG-1997 (first entry)

DE Banana polyphenol oxidase gene (clone BANPPO1).

KW primer: degenerate; PCR: polymerase chain reaction; amplify; banana;

KW lettuce: polyphenol oxidase; PPO; conserved region; copper binding site;

KW Cua; GEN3; GEN8; antisense construct; decrease; prevent; browning;

KW plant; Lactuca sativa; Musa sapientum; ss.

OS Musa sapientum.

FT Key Location/Qualifiers

FT cds 2..1825

FT /*tag= a

FT /note= "no start codon is given and the reading frame is

FT interrupted by a stop codon at positions 38-40"

PN WO9637617-A1.

PD 28-NOV-1996.

PF 22-MAY-1996; AU0310.

PR 23-MAY-1995; AU-003098.

PR 26-SEP-1995; AU-005600.

PA (CSIR) COMMONWEALTH SCI & IND RES ORG.

PI Robinson SP;

DR WPI: 97-033994/03.

DR P-PSDB: M09422.

PT DNA encoding banana or lettuce polyphenol oxidase - useful in

PT preventing browning in plants

PT Claim 14; Fig 2; 33pp; English.

CC The present sequence, BANPPO1 encodes banana polyphenol oxidase (PPO).

CC It was derived by combining the sequences from various clones (BP03,

CC BP07 and BP056) isolated using degenerate (T47685-87) and specific

CC primers (T47692-94). A number of clones were obtained from the banana

CC fruit cDNA using primers B25 and GEN7 (T47695-96). Most of these clones

CC were identical to BANPPO1 but one clone, designated BANPPO11, was found

CC to be distinctly different (see T47699). The primers were designed based

CC on known plant PPO DNA sequences in the conserved regions of the gene

CC which encode the copper binding sites, Cua and CUB. Antisense banana or

CC lettuce PPO constructs can be used to decrease the level of PPO activity

CC in a banana or lettuce plant. Decreasing the level of PPO is useful in

CC preventing browning in plants.

SQ Sequence 2078 BP; 489 A; 629 C; 561 G; 399 T;

Query Match 49.7%; Score 17.4; DB 1; Length 2078;

Best Local Similarity 77.8%; Pred. No. 48;

Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 5 AAGCTTGAGTACGTCACGACGACGCT 31

DB 1262 AAGCTTCGGTACGACGTCGACGACGCT 1288

RESULT 11

Q40945/c

ID 040945 standard; DNA: 5433 BP.

AC 040945;

DE 08-SEP-1993 (first entry)

DE Human DNA polymerase alpha catalytic DNA sequence (corrected).

KW DNA polymerase: alpha catalytic polypeptide; hybridisation; ss;

KW template-dependent enzymatic nucleic acid synthesis; frameshift.

OS Homo sapiens.

FT Key Location/Qualifiers

FT cds 17..4405

FT /*tag= a

FT 1523..1534

FT /*tag= b

FT /note= "mutation from known sequence"

FT 2525..2527

FT /*tag= c

FT /note= "mutation from known sequence"

PN WO9310139-A.

PD 27-MAY-1993.

PE 12-NOV-1992; U09457.

PR 15-NOV-1991; US-792600.

PA (STRD) UNIV LELAND STANFORD JUNIOR.

PI Copeland WC, Wang TSF;

DR WPI: 93-182482/22.

DR P-PSDB: R37508.

PT Purified nucleic acid for enzymatic synthesis of nucleic acid -

PT encodes human polymerase alpha catalytic polypeptide, and is

PT contained in baculovirus vector for enhanced prodn. in insect

PT cells

PS Claim 2; Fig 3; 84pp; English.

CC Human DNA polymerase alpha catalytic polypeptides were isolated from

CC a human epidermoid carcinoma cell line, and seven peptides, T9, T19,

CC T23, T24, T25, T264 and T265, were sequenced. Each peptide sequence

CC was used to design anti-sense probes which were used to probe human

CC KB cDNA libraries to recover positive colonies contg. fragments

CC encoding DNA polymerase polypeptides. One cDNA clone obtd. was

CC used to screen a human pre-B cell cDNA library in lambda gt10 to

CC obtain a full length clone. Initial attempts to functionally express

CC the clone for human DNA polymerase alpha catalytic subunit resulted

CC in truncated translation prods. Resequencing of five overlapping

CC clones in conjunction with in vitro translation analysis revealed

CC two frame shift mutations and two missense mutations in the two

CC previously isolated cDNA clones, E1-19 and E1-12, that contain the 5'

CC end of the cDNA sequence. The sequence was mutated to correct the

CC variations and a full length ORF of 1462 amino acids produced.

CC See also Q40921-49.

SQ Sequence 5433 BP; 1672 A; 1052 C; 1272 G; 1437 T;

Query Match 49.1%; Score 17.2; DB 1; Length 5433;

Best Local Similarity 73.3%; Pred. No. 68;

Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 5 AAGCTTGAGTACGTCACGACGACGCTACT 34

DB 4748 AAGCTTCGGTACGTCACGTCACGCTACT 4719

RESULT 12

T13635

ID T13635 standard; DNA: 133894 BP.

AC T13635;

DE 03-SEP-1996 (first entry)

DE ACNPV genomic DNA clone 6.

KW Autographa californica nuclear polyhedrosis virus clone 6;

KW disruption: non-essential gene; heterologous protein production;

KW expression vector; baculovirus; ss.

OS Autographa californica nuclear polyhedrosis virus clone 6.

PN WO9601320-A2.

PD 18-JAN-1996.

PE 30-JUN-1995; IB0578.

PR 04-JUL-1994; GB-013420.

PA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.

PI Ayres M, Bishop D, Possee R;

DR WPI: 96-087670/09.

DR GENBANK: L22858.

PT Autographa californica nuclear polyhedrosis virus complete genome

PT sequence - useful in the prodn. of vectors for enhanced

PT heterologous protein expression, such as interleukin(s),

PT interferon(s) and neurotoxin(s)

PS Disclosure: Page 90-186; 122pp; English.

CC The complete nucleotide sequence of the genome of clone 6 of the

CC baculovirus Autographa californica nuclear polyhedrosis virus (ACNPV)

CC has been determined. The sequence is taken from the Genbank record

CC L22858. The patent specification claims a polynucleotide selected from

CC open reading frames (ORFs 13, 20, 22-26, 28-30, 32, 38, 41-46, 50-60,

CC 62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130,

CC 140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by

CC the patentees. See T13636-731. Expression vectors confg. the complete

CC genomic sequence of ACNPV, with the exception that at least one non-

CC essential ORF is disrupted or replaced are useful for the synthesis of
CC heterologous proteins.
SQ Sequence 133894 BP; 39195 A; 27151 C; 27347 G; 40201 T;

Query Match 49.1%; Score 17.2; DB 1; Length 133894;
Best Local Similarity 73.3%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 TCTTAAGCTTGAGTACGACGACGACG 30
Db 10003 TGTAAAGTTTGGCGTGGAGTCTGTCGAAC 10032

RESULT 13
T13636//C
ID T13636 standard; DNA; 984 BP.
AC T13636;
DE 06-SEP-1996 (first entry)
DM ACNPV ORF 13, residues 10621-9683.
KM Autographa californica nuclear polyhedrosis virus clone 6;
KW disruption; non-essential gene; heterologous protein production;
KM expression vector; baculovirus; ss.
OS Autographa californica nuclear polyhedrosis virus clone 6.
FH Key Location/Qualifiers
FT cds
FT 1..984
FT /*tag- a
FT /number= ORF 13
FT /note= "corresponds to ACNPV nucleotides
FT complement (10621-9683)"

PN W09601320-A2.
PD 18-JAN-1996.
PE 30-JUN-1995; IB0578.
PR 04-JUL-1994; GB-013420
PA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
PI Ayres M, Bishop D, Possee R;
DR GENBANK; L22858.
PT Autographa californica nuclear polyhedrosis virus complete genome
PT sequence - useful in the prodn. of vectors for enhanced
PT heterologous protein expression, such as Interleukin(s),
PT Interferon(s) and neurotoxin(s)
PS Claim 1; Page 90-186; 122PP; English.
CC T13636-731 show open reading frames 13, 20, 22-30, 32, 38, 41-46, 50-60,
CC 63-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-127, 129-130, 140-
CC 146, 148-150, 152 and 154 from a total of 154 ORFs identified in the
CC Autographa californica nuclear polyhedrosis virus (ACNPV) clone 6. Each
CC gene is numbered according to its position in the virus genome beginning
CC at the left end of the linear map, and irrespective of its orientation.
CC The direction of transcription is relative to that of the polyhedrin
CC gene. Of the 154 ORFs identified it was found that some of the ORFs (ORFs
CC 2, 30, 32, 71, 86, 123, 126 and 127) are dispensable for virus
CC replication in cell culture or insect larvae. These genes can be deleted
CC from the genome to: (a) provide additional sites for inserting single or
CC multiple copies of foreign genes; and (b) to reduce the size of the virus
CC genome. The present sequence ORF 13 is on the complementary strand
CC relative to the polyhedrin gene.
SQ Sequence 984 BP; 314 A; 211 C; 227 G; 232 T;

Query Match 49.1%; Score 17.2; DB 1; Length 984;
Best Local Similarity 73.3%; Pred. No. 52;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 TCTTAAGCTTGAGTACGACGACGACG 30
Db 619 TGTAAAGTTTGGCGTGGAGTCTGTCGAAC 590

RESULT 14
V03349
ID V03349 standard; DNA; 1679 BP.
AC V03349;

DT 17-JUN-1998 (first entry)
DE Homo sapiens interleukin-12 p35 subunit gene promoter sequence.
KW Interleukin-12; IL-12; p35 subunit; promoter; disease; treatment;
KM cancer; viral infection; HIV; autoimmune; infectious; allergic; ds.
OS Homo sapiens.

FH Key Location/Qualifiers
FT misc_signal
FT 1352
FT /*tag- a
FT /note= "alternative transcription start site"
FT 1594
FT /*tag- b
FT /note= "alternative transcription start site"
FT 1625
FT misc_signal
FT 1625
FT /*tag- c
FT /note= "alternative transcription start site"
FT 1626
FT /*tag- d
FT /note= "alternative transcription start site"

PN EP-818534-A1.
PD 14-JAN-1998.
PE 12-JUL-1996; 201977.
PR 12-JUL-1996; EP-201977.
PA (JANC) JANSSEN PHARM NV.
PI De Chaftey de Courcelles DRGG, Roevens PWM, Van Dijk PUG;
DR DNA comprising human interleukin-12 p35 subunit gene promoter - and
PT comprising first intron of IL-12 p40 subunit gene, useful for
PT screening for compounds to treat e.g. cancers, allergies,
PT infections, etc
PS Claim 2; Page 12-13; 23PP; English.
CC The sequence is that of a human interleukin (IL)-12 p35 subunit gene
CC promoter sequence. Cells containing a construct in which a reporter
CC gene is placed under the control of the IL-12p35 and p40 subunit
CC gene expression regulating sequences, can be used to screen for
CC compounds which affect the production of IL-12 in cells, especially
CC human cells. These compounds especially non-proteinaceous compounds
CC different from lipopolysaccharide, can be used for the treatment of
CC a disease involving a too high or too low IL-12 production, e.g.
CC cancers, viral infections including HIV, immunodeficiencies,
CC autoimmune diseases, infectious diseases or allergic diseases.
CC The nucleic acid sequences can be used to make transgenic non-human
CC mammalian animals which expresses a reporter gene under their control.
SQ Sequence 1679 BP; 412 A; 408 C; 406 G; 452 T;

Query Match 49.1%; Score 17.2; DB 1; Length 1679;
Best Local Similarity 73.3%; Pred. No. 57;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 2 CTTAAGCTTGAGTACGACGACGACG 31
Db 1316 CTTAAGCTTGAGTACGACGACGACG 1345

RESULT 15
V74417
ID V74417 standard; DNA; 4854 BP.
AC V74417;
DE 16-MAR-1999 (first entry)
DM Staphylococcus aureus contig SEQ ID #106.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature
FT 1321..1380
FT /*tag- a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 04:26:29 ; Search time 727.51 Seconds

(without alignments)
164,915 Million cell updates/sec

Title: US-09-077-173a-4

Perfect score: 35

Sequence: 1 TCTTAAGCTGAGTCACCTACGACGACCTAGTT 35

Scoring table: IDENTITY_NUC

Searched: 4089388 seqs, 1713965092 residues

Database: EST.*

Word size: 0

Number of hits that pass the threshold: 8178776

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*

50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_gss1:*
70: gb_gss2:*
71: gb_gss3:*
72: gb_gss4:*
73: em_gss1:*
74: em_gss2:*
75: em_gss3:*
76: em_gss4:*
77: gb_gss5:*
78: gb_gss6:*
79: gb_gss7:*
80: gb_gss8:*
81: gb_gss9:*
82: em_gss5:*
83: em_gss6:*
84: em_gss7:*
85: em_gss8:*
86: em_gss9:*
87: em_gss10:*
88: em_gss11:*
89: gb_gss10:*
90: gb_gss11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21.2	60.6	403	26	W34668 mc32e10.r1
2	21.2	60.6	507	29	AA049196 m149c10.r
3	21.2	60.6	478	35	AA538294 v198e05.r
4	21.2	60.6	446	43	A1158502 ud25b03.r
5	20.2	58.3	260	39	AA895889 vy35f03.r
6	20.2	57.7	424	21	T59835
7	20.2	57.7	648	44	A1294097
8	20.2	57.7	453	81	AO605448
9	19.8	56.6	410	23	H32055
10	19.8	56.6	292	41	A1030851
11	19.8	56.6	317	43	A1178808
12	19.8	56.6	493	43	A1229779
13	19.8	56.6	543	44	A1317823
14	19.8	56.6	443	46	A1028963
15	19.8	56.6	564	46	A1409203
16	19.6	56.0	500	25	N57501
17	19.6	56.0	366	46	A1438032
18	19.6	56.0	351	46	A1440573
19	19.6	56.0	345	46	A1440718
20	19.6	56.0	419	46	A1441460
21	19.6	56.0	398	46	A1443140
22	19.6	56.0	616	61	A1855807

Query Match	60.68;	Score 21.2;	DB 35;	Length 478;
Best Local Similarity	76.58;	Pred. No. 12;		
Matches 26;	Conservative 0;	Mismatches 8;	Indels 0;	Gaps 0

Best Local Similarity 76.5%; Pred. No. 12;
Matches 26; Conservative 0; Mismatches

Db 126 TCTGAGGCTGTGATGGCATGGAGGAGCAAGCCAGT 93

LOCUS	446 bp	mRNA
DEFINITION	u025h03 r1 Soares 2NbmT Mus musculus	
RESULTS		
AI158502/c		

ACCESSION	A1158502	
NID	93686971	
VERSION	A1158502.1	GI:3686971
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE
AUTHORS
1 (bases 1 to 446)
Lutneria; Rodentia; Sciurognathini; Mur
Marra, M., Hillier, L., Allen, M., Bowl
Giesel, S., Kucaba, T., Lacy, M., Le, M.
Schell, J.

Thelsing, B., Wylle, T., Lennon, G., So
Waterston, R.
The WashU-HMII Mouse EST Project
Submitted (1006)

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63110
Tel: 314 286 1800

Fax: 314 266 1810
 Email: mouse@watson.wustl.edu
 This clone is available royalty-free
 IMC Consortium (info@imcge.org)
 MGI:920209
 Seq primer: -28m13 rev2 ET from Ameri
 High quality sequence shot: 446.

```

source
1. .446
location/Vaultiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

```

/clone_id="Soares_ZNDM1"
 /sex="male"
 /tissue_type="Thymus"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT733-Pac (Pharmacia) with a modified
 polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer (5',
 TGTTCCCAATCTGACGTGGAGCGCCCTTTTTTTTTTTTTTTTTT
 3'); double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT733 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through two
 rounds of normalization, and was constructed by Bento
 Soares and M.Fatima Bonaldo."

Query Match	60.68;	Score 21.2;	DB 43;	Length 446;
Best Local Similarity	76.58;	Pred. No. 12;		
Matches 26; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;

LOCUS	AI294097	648 bp	mRNA	EST	01-DEC-1998
DEFINITION	LP07395.5prlme LP Drosophila melanogaster larval-early pupal por2				
ACCESSION	Drosophila melanogaster cDNA clone LP07395 5prlme, mRNA sequence.				
NID	AI294097				
VERSION	93943504				
KEYWORDS	AI294097.1 GI:3943504				
SOURCE	EST.				
ORGANISM	fruit fly.				
REFERENCE	Drosophila melanogaster				
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
TITLE	1 (bases 1 to 648)				
JOURNAL	Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,				
COMMENT	Brakestein, P., Lewis, S. and Rubin, G.M. BDGP/HMT Drosophila EST Project Unpublished (1997) On Oct 17, 1997 this sequence version replaced gi:2518093.				
FEATURES					
SOURCE	Source				
	Contact: Harvey, D. G. M. Rubin-Molecular and Cell Biology University of California Berkeley 539 LSA, Berkeley, CA 94720-3200, USA Fax: 510 643 9947 Email: http://www.fruitfly.org/EST/estfruitfly.berkeley.edu hit genomic sequence AL031583 Plate: 73 row: H column: 11 High quality sequence stop: 546. Location/Qualifiers 1..648 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /map="22" /clone="LP07395" /clone_lib="LP Drosophila melanogaster larval-early pupal POR2" /sex="male and female" /dev_stage="larvae-pupae" /lab_host="DHS-alpha" /note="Organ: whole body; Vector: POR2; Site:1: ECORI; Site:2: XhoI; Sized fractionated cDNAs were directly ligated into POR2. Plasmid cDNA library."				
BASE COUNT	180 a 150 c 151 g 167 t				
ORIGIN					
	Query Match 57.7%; Score 20.2; DB 44; Length 648; Best Local Similarity 75.8%; Pred. No. 37; Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;				
OY	1 TCTTAGCTGTGAGTCACGTACGACAGCTAG 33				
Db	 501 TCTTAGCTGTGAGTCACGTACGATGAGCAGATG 533				
RESULT	8				
LOCUS	AO605448/c				
DEFINITION	AO605448 453 bp DNA GSS 10-JUN-1999				
ACCESSION	HS-2119-B2_D07_MK CIT Approved Human Genomic Sperm Library D Homo				
NID	sequence. genomic clone Plate-2119 Col-14 Row-H, genomic survey				
VERSION	AO605448 95065372				
KEYWORDS	AO605448.1 GI:5065372				
SOURCE	GSS.				
ORGANISM	human.				
	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 453) Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,				

TITLE	Shaker, R., Schmidt, S., Traicoff, R. and Hood, L.E.		
JOURNAL	Sequencing of a Characterized Clone Resource for Genomic		
COMMENT	Unpublished (1998)		
	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.husc.washington.edu Plate: 2119 Row: H Column: 14 Seq primer: M13 Reverse Class: BAC ends High quality sequence stop: 453. Location/Qualifiers 1. 453 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="Plate-2119 Col-14 Row-H" /clone_11b="CIT Approved Human Genomic Sperm Library D" /sex="Male" /note="Organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"		
BASE COUNT	111 a	95 c	89 g 155 t 3 others
ORIGIN			
	Query Match 57.7%; Score 20.2; DB 81; Length 453; Best Local Similarity 75.8%; Pred. No. 34; Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;		
QY	1 TCTTACGCTGGAGTCACGCTACGACCACTG 33 		
Db	56 TCTTACGCTGAGAGTCATGTTCCAGAAATGCGAG 24		
RESULT 9			
LOCUS	H32055 410 bp mRNA EST 02-APR-1998		
DEFINITION	EST106796 Rat PC-12 cells, untreated Rattus sp. cDNA clone RPBG19		
ACCESSION	H32055 5' end, mRNA sequence.		
NID	9977472		
VERSION	H32055.1 GI:977472		
KEYWORDS	EST.		
SOURCE	Rattus sp.		
ORGANISM	Rattus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	Lee, N.H., Weinstein, K.G., Kirkness, E.F., Earle-Hughes, J.A., Fuldner, R.A., Marmaras, S., Glodde, A., Gocayne, J.D., Adams, M.D., Kerlavage, A.R., Fraser, C.M. and Venter, J.C. Comparative expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995) 95396786		
TITLE			
JOURNAL			
MEDLINE			
COMMENT	Contact: Lee, NH ATCC The Institute for Genomic Research 9712, Medical Center Drive, Rockville, MD 20850, USA Tel: (301)-838-3529 Fax: (301)-838-0208 Email: nhlee@tigr.org For clone availability please contact the TIGR Database (tbdinfo@db.tigr.org) Seq primer: M13 Reverse.		


```

BASE COUNT      105 a      86 c      52 g      74 t
ORIGIN

Query Match
Best Local Similarity 56.6%; Score 19.8; DB 43; Length 317;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 5 AACCTGGAGTCACTGACGAGCAAGCTAGTT 35
    ||||| ||||| ||||| ||||| ||
Db 195 AACCTGGAGTCACTGACGAGCAAGCTAGTT 165

RESULT 12
AI229779/c      493 bp      mRNA      EST      20-JAN-1999
DEFINITION      EST226474 Normalised rat embryo, Bento Soares Rattus sp. cDNA clone
ACCESSION      AI229779
NID            93813666
VERSION        AI229779.1 GI:3813666
KEYWORDS       EST.
SOURCE         Rattus sp.
ORGANISM       Rattus sp.
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE      1 (bases 1 to 493)
                Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
                Kerlavage,A.R. and Adams,M.D.
                Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
                Gene Index
                Unpublished (1998)
JOURNAL        On Aug 21, 1998 this sequence version replaced.
COMMENT        Contact: Lee, NH
                AFCC
                The Institute for Genomic Research
                9712, Medical Center Drive, Rockville, MD 20850, USA
                Tel: (301)-838-3529
                Fax: (301)-838-0208
                Email: nhlee@tigr.org
                Seq primer: M13-21.
                Location/Qualifiers
                1..493
                /organism="Rattus sp."
                /db_xref="AFCC (1998):2040031"
                /db_xref="taxon:10118"
                /clone="REMCL85"
                /clone_1lb="Normalized rat embryo, Bento Soares"
                /dev_stage="embryo 8, 12, 18 dpc"
                /note="Vector: pT733pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT      157 a      116 c      113 g      107 t
ORIGIN

Query Match
Best Local Similarity 56.6%; Score 19.8; DB 43; Length 493;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 5 AACCTGGAGTCACTGACGAGCAAGCTAGTT 35
    ||||| ||||| ||||| ||||| ||
Db 194 AACCTGGAGTCACTGACGAGCAAGCTAGTT 164

RESULT 13
AI17823/c      543 bp      mRNA      EST      17-DEC-1998
LOCUS          EST234494 PC12 cells, untreated, pT733pac, TIGR Rattus sp. cDNA
DEFINITION      Clone RPPA31 3' end, mRNA sequence.
ACCESSION      AI17823
NID            94033090
VERSION        AI17823.1 GI:4033090
KEYWORDS       EST.

```

```

SOURCE          Rattus sp.
ORGANISM        Rattus sp.
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE      1 (bases 1 to 543)
                Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
                Kerlavage,A.R. and Adams,M.D.
                Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
                Gene Index
                Unpublished (1998)
JOURNAL        On Jan 19, 1998 this sequence version replaced gi:2151434.
COMMENT        Contact: Lee, NH
                AFCC
                The Institute for Genomic Research
                9712, Medical Center Drive, Rockville, MD 20850, USA
                Tel: (301)-838-3529
                Fax: (301)-838-0208
                Email: nhlee@tigr.org
                Seq primer: M13-21.
                Location/Qualifiers
                1..543
                /organism="Rattus sp."
                /db_xref="taxon:10118"
                /clone="RPPA31"
                /clone_1lb="PC12 cells, untreated, pT733pac, TIGR"
                /cell_line="PC12"
                /note="Vector: pT733pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT      165 a      154 c      107 g      117 t
ORIGIN

Query Match
Best Local Similarity 56.6%; Score 19.8; DB 44; Length 543;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 5 AACCTGGAGTCACTGACGAGCAAGCTAGTT 35
    ||||| ||||| ||||| ||||| ||
Db 190 AACCTGGAGTCACTGACGAGCAAGCTAGTT 160

RESULT 14
AI028963/c      443 bp      mRNA      EST      04-JUL-1999
LOCUS          UI-R-CO-10-f-11-0-UI.s1 UI-R-CO Rattus norvegicus cDNA clone
DEFINITION      UI-R-CO-10-f-11-0-UI 3', mRNA sequence.
ACCESSION      AI028963
NID            94295273
VERSION        AI028963.1 GI:4295273
KEYWORDS       EST.
SOURCE         Norway rat.
ORGANISM       Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE      1 (bases 1 to 443)
                Bonaldo,M.F., Lennon,G. and Soares,M.B.
                Normalization and subtraction: two approaches to facilitate gene
                discovery
                Genome Res. 6 (9), 791-806 (1996)
                97044477
                On Jun 22, 1998 this sequence version replaced gi:3246789.

```

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dr track served to identify it as a clone from the normalized
adult Placenta library. cDNA library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through

Mon Jan 3 08:07:27 2000

us-09-077-173a-4.rst

Page 9

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 07:18:38 ; Search time 601.47 Seconds

(without alignments)
248.818 Million cell updates/sec

Title: US-09-077-173A-3

Perfect score: 35

Sequence: 1 CAGATCTAGATCTATGTTCTACACCTCTTACGTC 35

Scoring table: IDENTITY_NUC

Searched: 780561 seqs, 2137953050 residues

Database : GenBank:*

Word size : 0

Number of hits that pass the threshold : 1561122

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vl:*
17: em_fun:*
18: em_hgt:*
19: em_hum1:*
20: em_hum2:*
21: em_in:*
22: em_om:*
23: em_or:*
24: em_ov:*
25: em_pat:*
26: em_ph:*
27: em_pl:*
28: em_ro:*
29: em_sts:*
30: em_sy:*
31: em_un:*
32: em_vl:*
33: gb_hgt1:*
34: gb_hgt2:*
35: gb_hgt3:*
36: gb_hgt4:*
37: gb_hgt5:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: gb_pr4:*
42: gb_hgt3:*
43: gb_hgt4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	100.0	35	5	A62987	A62987 Sequence 3
2	22	62.9	226345	41	AC005406	AC005406 , complet
3	21.6	61.7	70952	7	AB024028	AB024028 Arabidops
4	21.2	60.6	282610	1	RFX01	AJ35270 Ricketts
5	21.2	60.6	78921	7	AB028621	AB028621 Arabidops
6	21.2	60.6	213363	34	AC009499	AC009499 Homo sapi
7	20.8	59.4	94817	10	HS981L23	AL031686 Human DNA
8	20.8	59.4	1790	12	AF049125	AF049125 Mus muscu
9	20.6	58.9	235	13	G04095	G04095 human STS W
10	20.2	57.7	5701	7	SPCC1902	AL096592 S. pombe c
11	20.2	57.7	203006	11	HS1145L23	AL096699 Human DNA
12	20.2	57.7	147102	11	AC002476	AC002476 Human PAC
13	20.2	57.7	112359	11	HS060822	U60822 Human dylr
14	20.2	57.7	439	13	G21425	G21425 human STS W
15	20.2	57.7	163384	34	AC006962	AC006962 Homo sapi
16	20.2	57.7	12734	35	CEH16D19	AL031624 Caenorhab
17	20.2	57.7	95272	36	AC005298	AC005298 Drosophila
18	20	57.1	10167	2	MPAE000035	AE000035 Mycoplasma
19	20	57.1	184886	11	AC005358	AC005358 Homo sapi
20	20	57.1	21352	43	AC011659	AC011659 Homo sapi
21	19.8	56.6	72872	8	FL104	AF096370 Arabidops
22	19.8	56.6	1825	12	AB028143	AB028143 Mus muscu
23	19.8	56.6	33071	42	AC011030	AC011030 Homo sapi
24	19.8	56.6	217829	43	AC011361	AC011361 Homo sapi
25	19.6	56.0	74903	8	ATAC005312	AC005312 Arabidops
26	19.6	56.0	11457	11	AC003686	AC003686 Homo sapi
27	19.6	56.0	14676	33	AF029367	AF029367 Homo sapi
28	19.6	56.0	179285	33	CNS01DS6	AL121656 Homo sapi
29	19.6	56.0	220000	34	AC007126	AC007126 Homo sapi
30	19.6	56.0	133318	34	AC007770	AC007770 Drosophila
31	19.6	56.0	124646	34	AC007822	AC007822 Drosophila
32	19.4	55.4	12984	2	AE001122	AE001122 Borrelia
33	19.4	55.4	78600	7	SCDNC2	X79489 S. cerevisia
34	19.4	55.4	1425	7	SCROX3	X58300 S. cerevisia
35	19.4	55.4	2447	7	SCYBL092W	Z35853 S. cerevisia
36	19.4	55.4	1880	7	SCYBL093C	Z35854 S. cerevisia
37	19.4	55.4	167075	10	HS742J24	AL035448 Human DNA
38	19.4	55.4	143453	10	HSJ61G12	AL049797 Human DNA
39	19.4	55.4	1203	10	HUMDYSKW	M65036 Human dylr
40	19.4	55.4	38770	10	HUMDYSROP	M6524 Human dylr
41	19.4	55.4	7889	16	AF020905	AF020905 Common ch
42	19.4	55.4	241519	33	HS1012F16	AL080274 Homo sapi
43	19.4	55.4	91745	33	HSBG13C1	AL096802 Homo sapi
44	19.4	55.4	87319	34	AC004378	AC004378 Drosophila
45	19.4	55.4	37494	36	CELC04E12	AF078785 Caenorhab

ALIGNMENTS

RESULT	1	LOCUS	A62987	35 bp	DNA	PAT	12-MAR-1998
DEFINITION	Sequence 3 from Patent WO9719170.						
ACCESSION	A62987						
NID	93716859						
VERSION	A62987.1	GI:3716859					
KEYWORDS							
SOURCE	unidentified.						
ORGANISM	unclassified.						
REFERENCE	1 (bases 1 to 35)						
AUTHORS	Communi D., Piroton S., Parmentier M. and Boeynaems J.						
TITLE	RECEPTOR AND NUCLEIC ACID MOLECULE ENCODING SAID RECEPTOR						
JOURNAL	Patent: WO 97/19170-A 3 29-MAY-1997;						
FEATURES	EUROSCREEN S A (BE)						
source	Location/Qualifiers						
	1..35						
	/organism="unidentified"						

BASE COUNT 9 a /db_xref="taxon:32644"
ORIGIN 9 c 5 g 12 t

Query Match 100.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 8.3e-05;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGATCTAGTACTAGTCTACTACTCTACGTCG 35
Db 1 CAGATCTAGTACTAGTCTACTACTCTACGTCG 35

RESULT 2
AC005406 226345 bp DNA PRI 24-APR-1999
LOCUS AC005406
DEFINITION AC005406
ACCESSION 94580403
NID AC005406.2 GI:4580403
VERSION HTG.
KEYWORDS human.
SOURCE ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Homo sapiens 22q11 BAC Clone D128f11 in CES Region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 5 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 6 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 7 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 8 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 9 (bases 1 to 226345)

AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 10 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 11 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (26-NOV-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 12 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 13 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 14 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (09-APR-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 15 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (22-APR-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 16 (bases 1 to 226345)
AUTHORS Fu, Y., Pan, H., McDerimid, H. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (24-APR-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Apr 9, 1999 this sequence version replaced gi:3928116.
FEATURES
location/Qualifiers
1..226345
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 70721 a 47619 c 43447 g 64558 t
ORIGIN

Query Match 62.9%; Score 22; DB 41; Length 226345;
Best Local Similarity 83.3%; Pred. No. 6.3;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAGATCTAGTACTAGTCTACTACTCTA 30
Db 174495 CAGATCTAGTCTACTACTCAACACTGTA 174524

RESULT 3
AB024028/c 70952 bp DNA PLN 03-MAR-1999
LOCUS AB024028
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone: K1G2,
complete sequence.
ACCESSION AB024028
NID 94519187
VERSION AB024028.1 GI:4519187


```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
TITLE
FEATURES
source

HTG.
Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
clone:K1G2.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (sites)
Nakamura,Y.
Structural Analysis of Arabidopsis thaliana Chromosome 3. I
Unpublished (1998)
2 (bases 1 to 70952)
Nakamura,Y.
Direct Submission
Submitted (24-FEB-1999) to the DDBJ/EMBL/Genbank databases.
Yasuzaku Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
(E-mail:ynakamura@kazusa.or.jp, Tel:+81-438-52-3935,
Fax:+81-438-52-3934)
Location/Qualifiers
1..70952
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="K1G2"
/clone_lib="Mitsui P1"
BASE COUNT 23075 a 12412 c 12631 g 22834 t
ORIGIN

Query Match 61.7%; Score 21.6; DB 7; Length 70952;
Best Local Similarity 85.7%; Pred. No. 11;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAGACTAGATCTACTTCTTACTCTCT 28
Db 53707 CAGCTCTAGACTACTGTCTTAACCTAT 53680

RESULT 4
RPX01/c 282610 bp DNA BCT 11-NOV-1998
LOCUS Rickettsia prowazekii strain Madrid E, complete genome; segment
DEFINITION 1/4.
ACCESSION AJ235270 AJ235269
NID 93860572
VERSION AJ235270.1 GI:3860572
KEYWORDS complete genome.
SOURCE Rickettsia prowazekii.
ORGANISM Rickettsia prowazekii.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiae; Rickettsia.
1 (bases 1 to 282610)
Andersson,S.G., Zomorodipour,A., Andersson,J.O.,
Scheritz-Ponten,T., Alsmark,U.C., Podowski,R.M., Naslund,A.K.,
Eriksson,A.S., Winkler,H.H. and Kurland,C.G.
The genome sequence of Rickettsia prowazekii and the origin of
mitochondria
Nature 396 (6707), 133-140 (1998)
99039499
2 (bases 1 to 282610)
Andersson,S.G.E.
Direct Submission
Submitted (11-NOV-1998) S.G.E. Andersson,
Siv. Andersson@olbio.uu.se, Dept. of Molecular Biology, University
of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
Location/Qualifiers
1..282610
/organism="Rickettsia prowazekii"
/strain="Madrid E"
/db_xref="taxon:782"

gene
CDS
1..822
/gene="RP001"
1..822
/gene="RP001"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14477.1"
/db_xref="PID:e1342317"
/db_xref="PID:g3860573"
/db_xref="GI:3860573"
/translation="MTKLIIHLVSDSVGTAKAANSALAQFTSTKQKIHMPMINC
ELNEVLSTESKHGIVLTADQELKTLTFCYELKIPCISVIGKIIKESVSGI
ELKEQNYNKKERKEDYEDTLNADVAIRHDDGMINELSESDIILGSPRTSKTPSV
FLAVNGLKANIPIYVNCPEPDFIEKIDIDLVGLVINPRLIEIRARLNLQIEN
KSYTDNIQVRECIEVRKICNQNMWVIDVSTSIETALMRIYNNKNYHK"
906..1298
/gene="RP002"
906..1298
/gene="RP002"
/codon_start=1
/transl_table=11
/product="THIOREDOXIN (trxA)"
/protein_id="CA14475.1"
/db_xref="PID:e1342318"
/db_xref="PID:g3860574"
/db_xref="GI:3860574"
/translation="MSCVNEITLLEPDSNDINTQRIIMVNVVDSFRKNEVSLD
PVAVDPAERCGPCAKLIPIDETISKELDDKAVVLKMNIDENPKTPSEIGINSIPTIM
LFRNGQKDTKIGLQOKNSLLDWINKSI"
1309..2067
/gene="RP003"
1309..2067
/gene="RP003"
/codon_start=1
/transl_table=11
/product="O-ANTIGEN EXPORT SYSTEM ATP-BINDING PROTEIN REBE
(rfB)"
/protein_id="CA14476.1"
/db_xref="PID:e1342319"
/db_xref="PID:g3860575"
/db_xref="GI:3860575"
/translation="MLNHSKVEIETDVEGIDVKKRAQGLKHFLLKGVSLSPFI
PLNNINFSCEGEKIAFISNGSGSKSLKLAGIYPLKSGIVKVGHDIAIIMDVG
GFEQDQGRINKIMLYNNMLDKYSKTEKEIIDSELSGRIDLPKIVSGMSRL
APSVSIFQNPQIILLDEVFRAQDSYFTEKSLNLMKKFKNTPISTIVSHQETIKDNC
DRCILKDHITIDGTPSEIFKLYKOOSKEIKH"
2064..2840
/gene="RP004"
2064..2840
/gene="RP004"
/codon_start=1
/transl_table=11
/product="O-ANTIGEN EXPORT SYSTEM PERMEASE PROTEIN RFBA
(rfB)"
/protein_id="CA14477.1"
/db_xref="PID:e1342320"
/db_xref="PID:g3860576"
/db_xref="GI:3860576"
/translation="MIKYFFSKKYPRAIALLVKASIVRONKDSFLGSLNLIOPFIH
IVISYFGLRQREFVWNLVGGIPLMSFIYSIVSLICASSLVTRDQIIVKIVISKT
FPIIADISGLQVLTLCFSAMYPALFLPPEKRSQGIIMPLIILPVICVSGSIAY
AETPIYRIQPMNLVITIGYIVISPIVYPSLIPSKKIYFEFNFPLVIRPVQVLV
INSLDPMMLIKSVIAFTIVTISLYIRQFSKRVITYL"
2837..3724
/gene="RP005"
2837..3724
/gene="RP005"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14478.1"
/db_xref="PID:e1342321"

```

gene
CDS
/db_xref="PID:G3860577"
/translacion="WKLSFNLDNFNLDLTGLKKLDQITLDYLFADKSLAKHDLMLFPA
PLSTIDPDYSFKELKISPHDDFLAEFCJTSKEVYTSRLHAKOPDIIYECKRFVOR
VAVKRYPEKRIKIDFEDYVYLKTYMLIGTFNFSREFAKOIYIWOODEBSFSLDIAA
RFAARVSVSYSSNKNKMLFELNDILFNLQOKLDKKNLIDDKILKYOKNEVEDPD
YDSFENFDLEALNNSHCYICCHKRDKSCSGFDVIFPHFNAVISVDKIPRLSTMYMTG
GPLCKISPMYIKRQVLI"
3757. .5877
/gene="RP006"
3757. .5877
/gene="RP006"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14479.1"
/db_xref="PID:e1342322"
/db_xref="PID:G3860578"
/db_xref="GI:3860578"
/translacion="WVAATGHRICNDCKACITQKODPVNIPLESNIIEETLAPYG
LEIYILRBNPLNIYAPLPKPEPIYNIILVLGAGAGSLSLYLLRSGHWNTAIDGKLT
ITPSEFNHAKPYFWHEHYKNLSEIRIPGFGVAVMEYKALTSKNEGYEYKRLK
NKKYDVGALGPNITKFEQALDLSFDFHVAFCIGAGKPKLNTIENEAGVARSAPSELT
LONGAFCONSNTMIMRPIYIIGGLTSLDVATESIFYKKOVEEPKARYEIKDT
EEDKRIAEFETIAHAKLFEKAKNEELKVPKNGATYYIYGRLODSPARKLNEELI
YTLAGVNEKEMOPLRINIDKYGHVESVESITWLDNRTHKRYHINMCYVAPSSN
TYLITKYVIMAGIENNTQFDYDYSYFGDCNPPYEGSVKATSKAGEYVYKRLK
IDNVSDFKGLACFTOLDYLTSTRINKINILINDKTEFLIHSPLAKNEFGQFPEL
ONYSYDIKLEPVALSPADIDIEKGLTSFYIEYEGKSTCSTLSENEKVLVMPGPG
SLELPONKKIIEIIDSKRYNVLGLIEMENKKNYFAFYPIPKRKHLLTSDVIYINS
PIAEIOLDEIKFEGETELINVSNIQMCKGICGCTQYKVKOKYITRPSQNONN
VEIIDFKSLKTRLRNSIQEKRSN"
complement(6227. .7021)
/gene="RP007"
complement(6227. .7021)
/gene="RP007"
/codon_start=1
/transl_table=11
/product="ACYL-[ACYL-CARRIER-PROTEIN]- -UDP-N-
ACETYLGLUCOSAMINE O-ACETYLRANSFERASE (1pxa)"
/protein_id="CA14480.1"
/db_xref="PID:e1342323"
/db_xref="PID:G3860579"
/db_xref="GI:3860579"
/translacion="MSNSNIHTTALIAEGARGNKVKVGYPCIIIGPEVYLHDNVELKS
HVAIDIGIEGENTVITYPFASIGOPQITLKANERSSTIIGSNNTIREYTVVAGSKS
GMIRTVANNNLNEMGYHIDGCKIGNNLVYANSLAGHAKYGVDAIIGLSLVHGY
TRIGYSMIGGIPGADVIPGVASISRAVLEGINLTGMKRGKDFADSIITALNVE
EIIFGEGFVDRIRKVAEKYKNSIYTOIIDFLNODSSRAFCHFKK"
complement(7031. .7468)
/gene="RP008"
complement(7031. .7468)
/gene="RP008"
/codon_start=1
/transl_table=11
/product="(3R)-HYDROXYMYRISTOYL-[ACYL CARRIER PROTEIN]
DEHYDRATASE (fabz)"
/protein_id="CA14481.1"
/db_xref="PID:e1342324"
/db_xref="PID:G3860580"
/db_xref="GI:3860580"
/translacion="WIIATIEMLDIIHRYPFLVDRYLKDIPKNSIIGIKMYNVNCP
QTFGHPFAPVMPGVGLWETESMAQLAAILVAVASLSTYKKEVFLMSIENTFRFRIVQCG
DTMHASHVIDOORAVWKFSSKVVCEIAESKFTAMIKDKM"
complement(7659. .8699)
/gene="RP009"
complement(7659. .8699)
/gene="RP009"
/codon_start=1
/transl_table=11
/product="UDP-3-O-[3-HYDROXYMYRISTOYL] GLUCOSAMINE
N-ACETYLRANSFERASE (1pxd)"
/protein_id="CA14482.1"

	Query Match	60.6%	Score 21.2:	DB 1:	Length 282610;
	Best Local Similarity,	76.5%;	Pred. No. 13;		
	Matches 26;	Conservative 0;	Mismatches 8;	Indels 0;	Gaps 0;
QY	2 AGACTAGATCTATTGTTCTACACCTTCACTGC 35 Db 32913 AAATCTTAATAATATCTCTACTTTAAAGTC 32880				
RESULT 5	AB028621	78921 bp	DNA	PLN	11-JUN-1999
LOCUS	Arabidopsis thaliana genomic DNA, chromosome 3, pl clone:MU8,				
DEFINITION	complete sequence.				
ACCESSION	AB028621				
NID	g5041974				
VERSION	AB028621.1	GI:5041974			
KEYWORDS	HMG.				
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1				
ORGANISM	clone:MU8.				
	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core				
	eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;				
	Arabidopsi.				
REFERENCE	1 (sites)				
AUTHORS	Nakamura,Y.				
TITLE	Structural Analysis of Arabidopsis thaliana Chromosome 3. II				
JOURNAL	Unpublished (1999)				
REFERENCE	2 (bases 1 to 78921)				
AUTHORS	Nakamura,Y.				
TITLE	Direct Submision				
JOURNAL	Submitted (09-JUN-1999) to the DDBJ/EMBL/GenBank databases.				
	Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of				
	gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan				
	(E-mail:yinakam@kazusa.or.jp, Tel::+81-438-52-3935,				
	Fax:+81-438-52-3934)				
FEATURES	Location/Qualifiers				
SOURCE	1..78921				
	/organism="Arabidopsis thaliana"				
	/strain="Columbia"				
	/db_xref="taxon:3702"				
	/chromosome="3"				
	/clone="MU8"				
	/clone_lib="Mitsui P1"				
BASE COUNT	26493 a 14014 c 13288 g 25126 t				
ORIGIN					

Query Match 60.6%; Score 21.2; DB 7; Length 78921;
 Best Local Similarity 76.5%; Pred. No. 16;
 Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CAGATCTAGATACATCTTCTACACTCTTACGTC 34
 ||| ||||| ||||| ||| ||| |||
 Db 57900 CAAATCTAGATATATGTTCTACTATATGTG 57867

RESULT 6
 AC009499 212363 bp DNA HTG 24-AUG-1999
 LOCUS Homo sapiens clone NH0510D10, WORKING DRAFT SEQUENCE, 16 unordered
 DEFINITION pieces.
 AC009499
 NID 95762568
 VERSION AC009499.1 GI:5762568
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 212363)
 Watterston, R.H.
 The sequence of Homo sapiens clone
 Unpublished
 2 (bases 1 to 212363)
 Watterston, R.H.
 Direct Submission
 Submitted (24-AUG-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1987: contig of 1987 bp in length
 * 1988 2005: gap of unknown length
 * 2006 4154: contig of 2149 bp in length
 * 4155 4172: gap of unknown length
 * 4173 6140: contig of 1968 bp in length
 * 6141 6158: gap of unknown length
 * 6159 8980: contig of 2822 bp in length
 * 8981 8998: gap of unknown length
 * 8999 11264: contig of 2266 bp in length
 * 11265 11282: gap of unknown length
 * 11283 14408: contig of 3126 bp in length
 * 14409 14426: gap of unknown length
 * 14427 17873: contig of 3447 bp in length
 * 17874 17891: gap of unknown length
 * 17892 22734: contig of 4843 bp in length
 * 22735 22752: gap of unknown length
 * 22753 28660: contig of 5908 bp in length
 * 28661 28678: gap of unknown length
 * 28679 34714: contig of 6036 bp in length
 * 34715 34732: gap of unknown length
 * 34733 43061: contig of 8329 bp in length
 * 43062 43079: gap of unknown length
 * 43080 51186: contig of 8107 bp in length
 * 51187 51204: gap of unknown length
 * 51205 73417: contig of 22213 bp in length
 * 73418 73435: gap of unknown length
 * 73436 101632: contig of 28197 bp in length
 * 101633 101650: gap of unknown length
 * 101651 152551: contig of 50901 bp in length
 * 152552 152569: gap of unknown length
 * 152570 212363: contig of 59794 bp in length.

FEATURES
 Location/Qualifiers

source 1. 212363
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NH0510D10"

BASE COUNT 67114 a 41483 c 40358 g 63138 t 270 others
 ORIGIN

Query Match 60.6%; Score 21.2; DB 34; Length 212363;
 Best Local Similarity 76.5%; Pred. No. 14;
 Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CAGATCTAGATACATCTTCTACACTCTTACGTC 34
 ||| ||||| ||||| ||| ||| |||
 Db 189989 CAGCATGAGATACATCTTCTACACACTAGAGTC 190022

RESULT 7
 HS981L23/c
 LOCUS HS981L23
 DEFINITION Human DNA sequence from clone 981L23 on chromosome 20q12.1-13.2.
 Contains a Krueppel type zinc-finger protein pseudogene, a ZNF127
 pseudogene and a KRAB box type zinc-finger protein pseudogene.
 Contains ESTs, an STS, GSSs and a putative Cpg Island, complete
 sequence.
 ACCESSION AL031686
 NID 94581428
 VERSION AL031686.2 GI:4581428
 KEYWORDS HTG: Cpg Island; KRAB box; Krueppel; zinc-finger; ZNF127.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 94817)
 Babbage, A.
 Direct Submission
 Submitted (24-MAR-1999) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humqueye@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Apr 12, 1999 this sequence version replaced g1:4376020.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence is the entire insert of clone 981L23. This sequence
 has been finished according to sequence map criteria as follows. An
 attempt is made to resolve all sequencing problems, such as
 compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 This sequence was generated from part of bacterial clone contigs of
 human chromosome 20, constructed by the Sanger Centre Chromosome 20
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr20
 981L23 is from the library RPC15 constructed at the Roswell Park
 Cancer Institute by the group of Pieter de Jong. For further
 details see http://bacpac.med.buffalo.edu/VECTOR: pcyrac2.
 Location/Qualifiers
 1. 94817
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="981L23"
 /map="q12.1-13.2"
 /clone_lib="RPC15"
 repeat_region 83..133
 repeat_region 145..444
 repeat_region /note="L2 repeat: matches 2691..2741 of consensus"
 repeat_region /note="AluSg repeat: matches 1..300 of consensus"
 misc_feature complement(537..878)

```
/note="match: STS 294362"
repeat_region 814..960
/note="MIR repeat: matches 51..206 of consensus"
repeat_region 961..1238
/note="MER33 repeat: matches 1..320 of consensus"
repeat_region 1239..1311
/note="MIR repeat: matches 206..256 of consensus"
repeat_region 1384..1948
/note="L2 repeat: matches 2125..2750 of consensus"
misc_feature 1948..3122
/note="putative Cpg island"
repeat_region 1949..2251
/note="Alusx repeat: matches 1..301 of consensus"
repeat_region 2252..2390
/note="L2 repeat: matches 1998..2125 of consensus"
repeat_region 2326..2655
/note="65 copies 2 mer cc 59% conserved"
repeat_region 3599..3797
/note="MIR repeat: matches 16..227 of consensus"
repeat_region 3811..3899
/note="MIR repeat: matches 98..184 of consensus"
repeat_region 3929..4076
/note="MIR repeat: matches 67..212 of consensus"
repeat_region 4284..4374
/note="MER96 repeat: matches 1..93 of consensus"
repeat_region 4988..5172
/note="MIR repeat: matches 68..258 of consensus"
repeat_region 5260..5557
/note="Alusx repeat: matches 2..299 of consensus"
repeat_region 5597..6080
/note="L2 repeat: matches 2219..2728 of consensus"
repeat_region 6265..6571
/note="Alusx repeat: matches 1..299 of consensus"
repeat_region 6637..6799
/note="MIR repeat: matches 85..226 of consensus"
repeat_region 6894..7117
/note="MIR repeat: matches 20..251 of consensus"
repeat_region 7523..7626
/note="52 copies 2 mer gt 72% conserved"
repeat_region 7663..7964
/note="AluJb repeat: matches 1..301 of consensus"
repeat_region 8081..8269
/note="MIR repeat: matches 21..216 of consensus"
repeat_region 9420..9771
/note="L2 repeat: matches 2383..2710 of consensus"
repeat_region 9748..10112
/note="L2 repeat: matches 2068..2412 of consensus"
repeat_region 10521..10864
/note="MER1A1 repeat: matches 22..365 of consensus"
repeat_region 11136..11313
/note="L2R41 repeat: matches 604..766 of consensus"
repeat_region 11467..11659
/note="L2R41 repeat: matches 1..187 of consensus"
repeat_region 12018..12052
/note="MIR repeat: matches 109..143 of consensus"
repeat_region 12146..12269
/note="MIR repeat: matches 118..231 of consensus"
repeat_region 12658..12788
/note="MIR repeat: matches 107..259 of consensus"
repeat_region 12827..12855
/note="MIR repeat: matches 69..97 of consensus"
repeat_region 13169..13204
/note="18 copies 2 mer tg 97% conserved"
repeat_region 13826..14398
/note="HAL1 repeat: matches 211..756 of consensus"
repeat_region 14399..14620
/note="MER57A repeat: matches 1..211 of consensus"
repeat_region 14621..14925
/note="AluYb8 repeat: matches 1..311 of consensus"
repeat_region 14926..15117
/note="MER57A repeat: matches 211..434 of consensus"
repeat_region 15118..15199
/note="HAL1 repeat: matches 756..840 of consensus"

repeat_region 15226..15414
/note="MER20 repeat: matches 7..213 of consensus"
repeat_region 15771..16335
/note="L1MC repeat: matches 2372..2612 of consensus"
repeat_region 16355..16478
/note="62 copies 2 mer tt 69% conserved"
repeat_region 16386..16481
/note="3 copies 32 mer 83% conserved"
repeat_region 16486..16756
/note="Alusx repeat: matches 1..269 of consensus"
repeat_region 16948..17027
/note="40 copies 2 mer ta 65% conserved"
repeat_region 17120..17608
/note="L1M2 repeat: matches 5318..5825 of consensus"
repeat_region 17722..17833
/note="L1M4 repeat: matches 4940..5051 of consensus"
repeat_region 17834..18109
/note="MER33 repeat: matches 1..280 of consensus"
repeat_region 18110..18407
/note="AluJb repeat: matches 1..295 of consensus"
repeat_region 18408..18458
/note="MER33 repeat: matches 280..323 of consensus"
repeat_region 18459..18496
/note="L1M4 repeat: matches 4903..4940 of consensus"
repeat_region 18974..19254
/note="Alusx repeat: matches 1..296 of consensus"
repeat_region 20328..20596
/note="HERV repeat: matches 5075..5345 of consensus"
repeat_region 20938..20975
/note="19 copies 2 mer gt 95% conserved"
repeat_region 20979..21021
/note="MER3A repeat: matches 26..68 of consensus"
repeat_region 21030..21259
/note="Alusx repeat: matches 1..290 of consensus"
repeat_region 21261..21351
/note="MER3A repeat: matches 66..179 of consensus"
repeat_region 21493..22181
/note="L1M4 repeat: matches 2311..3034 of consensus"
repeat_region 22404..22828
/note="L2R16A repeat: matches 16..450 of consensus"
repeat_region 22833..23063
/note="L1MC4 repeat: matches 7741..7973 of consensus"
repeat_region 23220..23689
/note="TIGER1 repeat: matches 1..495 of consensus"
repeat_region 23690..24868
/note="L1M2 repeat: matches 5102..6308 of consensus"
repeat_region 24876..26802
/note="TIGER1 repeat: matches 473..2418 of consensus"
repeat_region 27799..27926
/note="4 copies 32 mer 81% conserved"
repeat_region 27799..27924
/note="63 copies 2 mer ac 80% conserved"
repeat_region 28166..28510
/note="HAL1 repeat: matches 403..752 of consensus"
repeat_region 28785..28914
/note="65 copies 2 mer aa 61% conserved"
repeat_region 29065..29165
/note="MT1J repeat: matches 403..516 of consensus"
repeat_region 29181..29282
/note="HERV repeat: matches 5211..5316 of consensus"
repeat_region 29358..29703
/note="HERV16 repeat: matches 2758..3096 of consensus"
repeat_region 29704..30328
/note="MT2D repeat: matches 1..553 of consensus"
repeat_region 30329..31860
/note="HERV16 repeat: matches 1688..2758 of consensus"
repeat_region 32014..32316
/note="Alusx repeat: matches 1..303 of consensus"
repeat_region 32346..32860
/note="MER34 repeat: matches 3..517 of consensus"
repeat_region 33110..33219
/note="L1ME1 repeat: matches 5563..5662 of consensus"
repeat_region 33225..33405
```

Query Match 59.4%; Score 20.8; DB 10; Length 94817;

Best Local Similarity 78.1%; Pred. No. 23; Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 ATCTAGATAGTATGTTCTACACTCTTACGTGC 35
 Db 70008 ATTTAGATATATGTTCTTCCATTTTAAATGC 69977

RESULT 8

AF049125

LOCUS AF049125 1790 bp mRNA ROD 06-MAR-1998
 DEFINITION Mus musculus talpoxin-associated calcium binding protein 49 mRNA,
 complete cds.

ACCESSION AF049125
 NID AF049125
 VERSION AF049125.1 GI:2935465

KEYWORDS

SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1790)
 Dadds, D., Schlimgen, A.K., Lu, S.Y. and Perin, M.S.
 Novel reticular calcium binding protein is purified on talpoxin
 columns

JOURNAL J. Neurochem. 64 (5), 2339-2344 (1995)
 MEDLINE 95239201
 REFERENCE 2 (bases 1 to 1790)
 AUTHORS Perin, M.S.

TITLE Mouse talpoxin-associated calcium binding protein 49 (TCBP49)
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1790)
 AUTHORS Perin, M.S.

TITLE Direct Submission
 JOURNAL Submitted (17-FEB-1998) Neurosciences, Cleveland Clinic Foundation,
 9500 Euclid Avenue, Cleveland, OH 44195, USA

FEATURES
 source location/Qualifiers
 1. 1790
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 194..1159
 /note="TCBP49; contains 6 EF hand domains"
 /codon_start=1
 /product="talpoxin-associated calcium binding protein 49"
 /protein_id="AAC05132.1"
 /db_xref="pid:g2935466"
 /db_xref="gi:2935466"
 /translation="MRLGPRPALGLLPLLLYAAGAASKEELHYPOGHRADYR
 EALLGVQEDVDYVKLGHEQQRLOSLIKKIDSDGFLTELELSQWIMSPKHYAM
 QEAKQOFEYDKNSDDPVTWDEYNIDWDYIDPEMNLADDTGEGSFRQHLIKRK
 REKANOQSGPGLSEEFIAFPEPEVDVMTFEVIOALEPHDKNDGVSLEEFIGD
 YRRDPANDEPMIYERKRFYNDYDKNDGRIDPOELISWVPNNQGLAQEVALHLI
 DEMDLSNDKSLSEELLENODLFVSDYDGRQLHDYFYHDEL"

BASE COUNT 490 a 384 c 449 g 467 t

ORIGIN
 Query Match 59.4%; Score 20.8; DB 12; Length 1790;
 Best Local Similarity 78.1%; Pred. No. 43;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 AGATCTAGATAGTATGTTCTACACTCTTACGT 33
 Db 1226 AGATATAGATGTTGTTCTTACACTCTTAAGT 1257

RESULT 9

G04095

LOCUS G04095 235 bp DNA STS 19-OCT-1995
 DEFINITION human STS WI-2864.
 ACCESSION G04095

REFERENCE 1 Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1790)
 Dadds, D., Schlimgen, A.K., Lu, S.Y. and Perin, M.S.
 Novel reticular calcium binding protein is purified on talpoxin
 columns

JOURNAL J. Neurochem. 64 (5), 2339-2344 (1995)
 MEDLINE 95239201
 REFERENCE 2 (bases 1 to 1790)
 AUTHORS Perin, M.S.

TITLE Mouse talpoxin-associated calcium binding protein 49 (TCBP49)
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1790)
 AUTHORS Perin, M.S.

TITLE Direct Submission
 JOURNAL Submitted (17-FEB-1998) Neurosciences, Cleveland Clinic Foundation,
 9500 Euclid Avenue, Cleveland, OH 44195, USA

FEATURES
 source location/Qualifiers
 1. 1790
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 194..1159
 /note="TCBP49; contains 6 EF hand domains"
 /codon_start=1
 /product="talpoxin-associated calcium binding protein 49"
 /protein_id="AAC05132.1"
 /db_xref="pid:g2935466"
 /db_xref="gi:2935466"
 /translation="MRLGPRPALGLLPLLLYAAGAASKEELHYPOGHRADYR
 EALLGVQEDVDYVKLGHEQQRLOSLIKKIDSDGFLTELELSQWIMSPKHYAM
 QEAKQOFEYDKNSDDPVTWDEYNIDWDYIDPEMNLADDTGEGSFRQHLIKRK
 REKANOQSGPGLSEEFIAFPEPEVDVMTFEVIOALEPHDKNDGVSLEEFIGD
 YRRDPANDEPMIYERKRFYNDYDKNDGRIDPOELISWVPNNQGLAQEVALHLI
 DEMDLSNDKSLSEELLENODLFVSDYDGRQLHDYFYHDEL"

BASE COUNT 490 a 384 c 449 g 467 t

ORIGIN
 Query Match 59.4%; Score 20.8; DB 12; Length 1790;
 Best Local Similarity 78.1%; Pred. No. 43;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 AGATCTAGATAGTATGTTCTACACTCTTACGT 33
 Db 1226 AGATATAGATGTTGTTCTTACACTCTTAAGT 1257

RESULT 10

G04095

LOCUS G04095 235 bp DNA STS 19-OCT-1995
 DEFINITION human STS WI-2864.
 ACCESSION G04095

REFERENCE 1 Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1790)
 Dadds, D., Schlimgen, A.K., Lu, S.Y. and Perin, M.S.
 Novel reticular calcium binding protein is purified on talpoxin
 columns

JOURNAL J. Neurochem. 64 (5), 2339-2344 (1995)
 MEDLINE 95239201
 REFERENCE 2 (bases 1 to 1790)
 AUTHORS Perin, M.S.

TITLE Mouse talpoxin-associated calcium binding protein 49 (TCBP49)
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1790)
 AUTHORS Perin, M.S.

NID 9721053
 VERSION G04095.1 GI:721053
 KEYWORDS STS sequence; primer; sequence tagged site.
 SOURCE human Random genome wide STS created from sheared whole human DNA.
 ORGANISM Homo sapiens

REFERENCE 1 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;
 Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
 Catarrhini; Homiidae; Homo.
 1 (bases 1 to 235)
 Hudson, T.

JOURNAL Whitehead Institute/MIT Center for Genome Research; Random Genome
 WIDE STS
 Unpublished (1995)
 REFERENCE 2 (bases 1 to 235)
 AUTHORS Hudson, T.

TITLE Whitehead Institute/MIT Center for Genome Research; Physically
 mapped STS
 JOURNAL Unpublished (1995)

COMMENT Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu

Primer A: GAGATGCCACCCCTCTC
 Primer B: AATGCTCCCTGTGTGATG
 STS size: 181
 PCR Profile:

Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
 Thermal Cycler:

Protocol:
 Template: 10 ng
 Primer: each 5 pm
 dNTPs: each 4 mM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul

Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3

FEATURES
 source location/Qualifiers
 1. 235
 /organism="Homo sapiens"
 STS
 primer_bind 55..72
 primer_bind complement(217..235)
 BASE COUNT 65 a 59 c 61 g 47 t 3 others

ORIGIN
 Query Match 58.9%; Score 20.6; DB 13; Length 235;
 Best Local Similarity 74.3%; Pred. No. 72;
 Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CAGATCTAGATAGTATGTTCTACACTCTTACGTGC 35
 Db 55 CAGATCAAGATGCTTTCTTCTAAGCTCTTCAAGTGC 21

RESULT 10
 LOCUS SPC1902 5701 bp DNA PLN 25-MAR-1999
 DEFINITION S.pombe chromosome III cosmid c1902.

ACCESSION NID	VERSION	KEYWORDS	SOURCE ORGANISM	REFERENCE AUTHORS TITLE	JOURNAL
AL096699 66015529					
AL096699.11	GI:6015529	HTG. human.	Homo sapiens Eukaryota; Metazoa; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 203006)	Pearce, A. Direct Submission Submitted (08-OCT-1999)	Sanger Centre, Hinxton, Cambridgeshire,

COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP. Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence is the entire insert of clone 1145123. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/chrX1145123> is from the library RPCR-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>

Location/Qualifiers

1..203006

source

```
source 1..203006
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="x"
/clonetype="1145123"
/map="p21.1-21.3"
/clone_id="RC1-5"
BASE COUNT 67748 a 37548 c 35939 g 61771 t
ORIGIN
```

Query Match	57.7%	Score 20.2	DB 10	Length 203006
Best Local Similarity	75.8%	Pred. No. 37		
Matches 25; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;

```

QY      1 CAGATCTAGATACTATGTTCTACACTCTTACGT 33
          ||| || ||| ||| ||| ||| ||| |||
Db 106668 CAGCTCAAGAGCCATGCTCTTACTCTTACAT 106700

```

RESULT 12
AC002476/c

LOCUS	AC002476	147102 bp	DNA	PRI	21-AUG-1997
DEFINITION	Human PAC clone DJ318C15 from Xq23, complete sequence.				
ACCESSION	AC002476				

NID	92340101	GI:2340101
VERSION	AC002476.1	
KEYWORDS	HTG.	

SOURCE ORGANISM	human . Homo sapiens
-----------------	-------------------------

REFERENCE
1 (bases 1 to 147102)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Graves, T and Ozersky, P.
 TITLE The sequence of H. sapiens PAC clone D3J16C15
 JOURNAL unpublished (1997)
 REFERENCE 2 (bases 1 to 447102)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-1997) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT SUBMITTED BY:

Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
<http://genome.wustl.edu/gsc>
<mailto:saplens@watson.wustl.edu>

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by David Bentley's chromosome X mapping group at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

SOURCE INFORMATION:
This clone was derived from human PAC library RCE1-1 prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. For further details, see <http://pacpac.med.buffalo.edu/>. The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>).
VECTOR: pCIPAC2

NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of H_DJ318C15.
The actual end is at 147102 of H_DJ318C15. The orientation of this
clone is unknown.

This clone contains STS's WI-13810 (NID:g1343514) and WI-9475 (NID:g860548).

FEATURES
source

```
misc_feature 1..332 /note="match to EST N24741 (NID:g1138891) yx56c02.s1"
misc_feature 1..440 /note="match to EST N31640 (NID:g1152039) yw84g03.s1"
misc_feature 1..569 /note="match to EST AA207254 (NID:g1802747) zq82e06.s1"
misc_feature /note="match to EST H23858 (NID:g893553) yn71f07.r1"
misc_feature /note="match to EST H23858 (NID:g893553) yn71f07.r1"
misc_feature /note="match to EST 233445 (NID:g486641)"
misc_feature /note="match to EST 233445 (NID:g486641)"
misc_feature /note="match to EST R23654 (NID:g778542) yh35g02.r1"
misc_feature 395..811
```

misc_feature	complement(432..753)	/note="match to EST R23654 (NID:g778542) yh35g02.r1"
misc_feature	/note="match to EST R5832 (NID:g778542) yh35g02.r1"	complement(481..958)
misc_feature	/note="match to EST NA2209 (NID:g1166240) yw71e02.r1"	complement(692..1162)
misc_feature	/note="match to EST AA209339 (NID:g1807353) zq82e06.r1"	complement(1229..1333)
misc_feature	/note="match to EST H09881 (NID:g874703) ym05b04.r1"	complement(1336..1393)
misc_feature	/note="match to EST R25050 (NID:g779938) yq42a06.r1"	complement(1418..1676)
misc_feature	/note="match to EST R25050 (NID:g779938) yq42a06.r1"	complement(1418..1676)
misc_feature	/note="match to EST H09881 (NID:g874703) ym05b04.r1"	complement(1658..1676)
repeat_region	/note="match to EST AA148755 (NID:g1718917) z009c05.r1"	1677..1716
misc_feature	/rpt_family="L1"	complement(1717..1972)
repeat_region	/note="match to EST AA148755 (NID:g1718917) z009c05.r1"	complement(2097..2120)
repeat_region	/rpt_family="L1"	2478..2513
repeat_region	/rpt_family="L1"	2894..3036
repeat_region	/rpt_family="L1"	complement(2897..3017)
misc_feature	/rpt_family="L1"	4778..5126
misc_feature	/note="match to EST R31677 (NID:g787520) yh63c10.s1"	4779..5302
misc_feature	/note="match to EST AA085744 (NID:g1629216) zk71b09.s1"	4779..4982
misc_feature	/note="match to EST F01897 (NID:g645454)"	complement(4779..5124)
misc_feature	/note="match to EST R65700 (NID:g838338) y116f12.r1"	complement(4779..4878)
misc_feature	/note="match to EST C21402 (NID:g1622512)"	complement(4790..4985)
misc_feature	/note="match to EST D58313 (NID:g964335)"	complement(5072..5302)
misc_feature	/note="match to EST H48596 (NID:g988436) yq80a02.r1"	5100..5194
misc_feature	/note="match to EST R31677 (NID:g787520) yh63c10.s1"	5101..5443
misc_feature	/note="match to EST R97806 (NID:g983466) yq60d03.s1"	complement(5132..5302)
misc_feature	/note="match to EST AA125797 (NID:g1687768) z122d02.r1"	complement(5161..43007)
gene	/gene="LAMP2"	complement(5161..5300,15568..15732,16437..16500,20143..20265,21679..21863,22808..22966,29195..29408,30489..30607,42844..43007)"
misc_feature	/gene="LAMP2"	complement(5338..6010)
misc_feature	/note="match to EST R97855 (NID:g983515) yq60d03.r1"	complement(5797..6213)
misc_feature	/gene="LAMP2"	complement(5797..6213)

Query Match	Best Local Similarity	75.8%	Score 20.2	DB 11	Length 147102
Matches	25	Conservative	0	Mismatches	8
				Indels	0
				Gaps	0
Qy	2	AGATCTAGTACTATGTTCTACACCTTACGCG	34		
Db	54603	AGCTCTAGAAATATCTCTCACCCTACCCGCG	54571		
LOCUS	HSU60822/c				
DEFINITION	Human dystrophin (DMD) gene, exons 7, 8 and 9, and partial cds.				
ACCESSION	U60822.1				
VERSION	92275172				
KEYWORDS	U60822.1 GI:2275172				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	McNaughton, J.C., Broom, J.E., Hill, D.F., Jones, W.A., Marshall, C.J., Renwick, N.M., Stockwell, P.A. and Petersen, G.B.				
TITLE	A cluster of transposon-like repetitive sequences in intron 7 of the human dystrophin gene				
JOURNAL	J. Mol. Biol. 232 (1), 314-321 (1993)				
MEDLINE	93323127				


```

REFERENCE      2 (bases 92090 to 94459)
AUTHORS        McNaughton,J.C., Marshail,C.J., Broom,T.E., Hughes,G., Jones,W.A.,
               Stockwell,P.A. and Petersen,G.B.
               Phylogenetic relationships among transposon-like elements in human
               and primate DNA
TITLE          J. Mol. Evol. 40 (2), 127-135 (1995)
JOURNAL        95214090
MEDLINE        3 (bases 1 to 112359)
REFERENCE      McNaughton,J.C., Hughes,G., Jones,W.A., Stockwell,P.A., Klamut,H.J
AUTHORS        and Petersen,G.B.
TITLE          The evolution of an intron: analysis of a long, deletion-prone
JOURNAL        intron in the human dystrophin gene
MEDLINE        Genomics 40 (2), 294-304 (1997)
REFERENCE      4 (bases 1 to 112359)
AUTHORS        McNaughton,J.C.
TITLE          Direct Submission
JOURNAL        Submitted (14-JUN-1996) Department of Biochemistry, University of
               Otago, Box 56, Dunedin, New Zealand
REFERENCE      5 (bases 1 to 112359)
AUTHORS        McNaughton,J.C.
TITLE          Direct Submission
JOURNAL        Submitted (23-JUL-1997) Department of Biochemistry, University of
               Otago, Box 56, Dunedin, New Zealand
REMARK         Sequence update by Submitter
COMMENT        On Jul 23, 1997 this sequence version replaced gi:1845599.
FEATURES       location/Qualifiers
               1..112359
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="X"
                /map="Xp21.2"
                /clone="cosmids XJC5, Tcoos5, P31"
                <5..112359
                 /gene="DMD"
                 /number=5
                 /rpt_family="LINE2"
                 /protein_id="AAC31631.1"
                 /cd_xref="PID:g1845600"
                 /db_xref="GI:1845600"
                 /translation="PDLFDWNSVVCOSATQRLERHAENIARYOIGIEKILIDPEVDYDT
                 YPKRSILMTITSLFQVLPOOVSTEAIOEVMLEPRPKVTYEENRFLQHQRHSQQOITTTT
                 VSLAQGERTSSRPFRKSAITTOAIVTTSDPTSRPPSQ"
                 1144..110717
                  /gene="DMD"
                  /number=7
                  complement(2901..2941)
                   /gene="DMD"
                   /rpt_family="L1MA2"
                   complement(3288..6448)
                    /gene="DMD"
                    /rpt_family="L1"
                    /gene="DMD"
                    /rpt_family="L1"
                    6545..10074
                     /gene="DMD"
                     /rpt_family="L1"

```

misc_feature	10324 /gene="DMD" /note="c(x:1) breakpoint noted in sequence with EMBL accession number Z21689"
misc_feature	10606..10648 /gene="DMD" /note="deleted in sequence with EMBL accession number Z21689 in association with t(x:1)"
repeat_region	13351..13495 /gene="DMD" /rpt_family="Alu-J (FRAM)" complement(13886..16015)
repeat_region	/gene="DMD" /rpt_family="L1MA5" complement(16016..16327)
repeat_region	/gene="DMD" /rpt_family="Alu-Ya" complement(16328..19040)
repeat_region	/gene="DMD" /rpt_family="L1" 19041..19302
repeat_region	/gene="DMD" complement(22862..23161)
repeat_region	/gene="DMD" /rpt_family="MER4" 23349..23422
satellite	/gene="DMD" 24514..24547
satellite	/gene="DMD" complement(26830..27016)
repeat_region	/gene="DMD" /rpt_family="MER2" complement(27017..27143)
repeat_region	/gene="DMD" /rpt_family="L1PA15" 27230..27689
satellite	/gene="DMD" complement(27690..27945)
repeat_region	/gene="DMD" /rpt_family="L1MA2" complement(29491..29791)
repeat_region	/gene="DMD" /rpt_family="Alu-Ya" 30071..30119
repeat_region	/gene="DMD" complement(30756..30916)
repeat_region	/gene="DMD" /rpt_family="L1" 30942..31113
repeat_region	/gene="DMD" /rpt_family="L1" 31689..32181
repeat_region	/gene="DMD" /rpt_family="MLTID" 33905..36231
repeat_region	/gene="DMD" /rpt_family="THE-1" 36934..37375
repeat_region	/gene="DMD" /rpt_family="L1MA5" 38036..38212
repeat_region	/gene="DMD" /rpt_family="MIR" 39126..39147
satellite	/gene="DMD" 41174..44154
repeat_region	/gene="DMD" /rpt_family="THE-1 (LTR)" complement(43736..43934)
repeat_region	/gene="DMD" /rpt_family="L1MA5" 44014..45070
repeat_region	/gene="DMD"

```

repeat_region /rpt_family="L1"
                complement(45071..45483)
                /gene="DMD"
repeat_region /rpt_family="THE-1 (LTR)"
                45484..46342
                /gene="DMD"
repeat_region /rpt_family="L1"
                complement(48450..48777)
                /gene="DMD"
repeat_region /rpt_family="Alu-Sc"
                complement(50028..50145)
                /gene="DMD"
repeat_region /rpt_family="MIR2"
                complement(51861..51959)
                /gene="DMD"
repeat_region /rpt_family="MIR"
                52678..52692
                /gene="DMD"
satellite      /note="(TTC)5"
                complement(52837..53199)
                /gene="DMD"
repeat_region /rpt_family="THE-1 (LTR)"
                53051..54002
                /gene="DMD"
misc_feature   /note="Inverted in sequence with GenBank accession number
                M18024 in association with t(x:21)(p21:p12)"
                53342..53583
                /gene="DMD"
repeat_region /rpt_family="L1MA9"
                55930
                /gene="DMD"
misc_feature   /note="t(x:21)(p21:p12) breakpoint noted in sequence with
                GenBank accession number M18024"
                56250..56498
                /gene="DMD"
repeat_region /rpt_family="MIR"

Query Match      57.7%: Score 20.2; DB 11; Length 112359;
Best Local Similarity 75.8%: Pred. No. 40;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CAGATCTAGTACTAGTCTTACACTCTTACGT 33
    ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38060 CAGCTCAGAGCCTATGCTTACTCTTACAT 38028

RESULT 14
G21425 439 bp DNA 31-MAY-1996
LOCUS human STS WI-13400, sequence tagged site.
DEFINITION G21425
ACCESSION G21425
NID 91341751
VERSION G21425.1 GI:1341751
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STS derived from sequences in dbEST and the Unigene
        collection.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 439)
AUTHORS Hudson,T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
        Mapped STS
JOURNAL Unpublished (1995)
COMMENT Contact: Thomas Hudson
        Whitehead Institute/MIT Center for Genome Research
        Whitehead Institute for Biomedical Research
        9 Cambridge Center, Cambridge MA 02142 USA
        Tel: 617 252 1900
        Fax: 617 252 1902
        Email: thudson@genome.wi.mit.edu

```

```

Primer A: TCGTTGATTCATCTTCATGTAGCA
Primer B: AAAGGTTTCAGAGAAATTATCTCA
STS size: 129
PCR Profile:
  Presoak:
  Denaturation:
  Annealing: 56 degrees C
  Polymerization:
  PCR Cycles: 35
  Thermal Cycler:
Protocol:
  Template: 10 ng
  Primer: each 5 pm
  dNTPs: each 4 mM
  Tag Polymerase: 0.025 units/ul
  Total Vol: 20 ul
Buffer:
  MgCl2: 1.5 mM
  KCl: 50 mM
  Tris-HCL: 10 mM
  pH: 9.3

FEATURES
    source          Derived from dbEST (genbank accession R496688).
    location/Qualifiers
        1..439
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /map="403.0 CR from top of Chr12 linkage group"
    STS
        primer_bind 44..68
        primer_bind 44..172
        primer_bind complement(148..172)
    BASE COUNT 135 a 67 c 82 g 149 t 6 others
    ORIGIN

Query Match      57.7%: Score 20.2; DB 13; Length 439;
Best Local Similarity 73.5%: Pred. No. 96;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CAGATCTAGTACTAGTCTTACACTCTTACGT 34
    ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 CATATGTATATAGTACTTGTACACTGNCACATG 236

RESULT 15
AC006962 163384 bp DNA 05-MAR-1999
LOCUS Homo sapiens clone DnA0052J04, WORKING DRAFT SEQUENCE, 3 unordered
        pieces.
ACCESSION AC006962
NID 94337300
VERSION AC006962.1 GI:4337300
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163384)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
        Unpublished
JOURNAL 2 (bases 1 to 163384)
REFERENCE Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1999) Genome Sequencing Center, Washington
        University School of Medicine, 4444 Forest Park Parkway, St. Louis,
        MO 63108, USA
COMMENT * NOTE: This is a 'working draft' sequence. It currently
        * consists of 3 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.

```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 29, 1999, 17:02:53 ; Search time 12.66 Seconds

(without alignments)
1359.937 Million cell updates/sec

Title: US-09-077-173a-2

Perfect score: 1946

Sequence: 1 MASTESSLLRLSLGLSPGSS.....CRMATPDSSCSTPRADR 365

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database: PIR_62.*

Word size: 0

Number of hits that pass the threshold: 142080

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1940	99.7	365	2	G protein-coupled
2	962.5	49.5	373	2	ATP receptor P2u -
3	934	48.0	375	2	P-2u nucleotide re
4	641.5	33.0	373	2	P2Y1 receptor G-pr
5	630	32.4	372	2	S54253
6	628	32.3	362	2	G protein-coupled
7	621	31.9	373	2	G protein-coupled
8	588	30.2	328	2	P2Y receptor - boy
9	586	30.1	326	2	a novel G protein-
10	586	30.1	328	2	P2 purinoceptor -
11	501.5	25.8	308	2	P216 receptor - hu
12	479.5	24.6	344	2	G protein-coupled
13	450.5	23.2	370	2	Inton 17 purinerg
14	423	21.7	399	2	hepatic P2Y5-
15	411.5	21.1	420	2	proteinase activat
16	403	20.7	397	2	thrombin receptor
17	403	20.7	397	2	proteinase-activat
18	403	20.7	397	2	proteinase-activat
19	400	20.6	372	2	proteinase-activat
20	397.5	20.4	425	2	delta oploid recep
21	388.5	20.0	427	2	thrombin receptor
22	388	19.9	432	2	alpha-thrombin rec
23	385	19.8	372	2	thrombin receptor
24	384	19.7	372	2	delta oploid recep
25	383	19.7	361	2	delta opiate recep
26	382.5	19.7	362	2	G protein-coupled
27	381.5	19.6	398	2	angiotensin II rec
28	378.5	19.5	400	2	mu oploid receptor
29	378.5	19.5	400	2	opiate receptor mu
30	377	19.4	372	2	oploid receptor mu
31	376.5	19.3	398	2	delta oploid recep
32	371.5	19.1	342	2	mu-oploid receptor
33	371.5	19.1	398	2	platelet-activatin
34	370.5	19.0	359	2	mu oploid receptor
35	369.5	19.0	359	2	G protein-coupled
			2		angiotensin II rec

36	365	18.8	363	2	I57955	somatostatin recep
37	364.5	18.7	359	2	JC2134	angiotensin II rec
38	364	18.7	364	2	JN0763	somatostatin recep
39	363	18.7	359	2	A42656	angiotensin II rec
40	362	18.6	362	2	S68207	G protein-coupled
41	361	18.6	354	2	I53033	G protein-coupled
42	360.5	18.5	359	2	S44425	angiotensin II rec
43	360	18.5	359	2	JC1104	angiotensin II rec
44	359	18.4	352	2	A57641	G protein-coupled
45	358.5	18.4	359	2	S15403	angiotensin II rec

ALIGNMENTS

RESULT 1
S68679
G protein-coupled receptor - human
C:Species: Homo sapiens (man)
C>Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 24-Sep-1999
C:Accession: S68679
R:Stam, N.J.; Klomp, J.; van de Heuvel, M.; Olijve, W.
FEBS Lett. 384, 260-264, 1996
A:Title: Molecular cloning and characterization of a novel orphan receptor (P(2P)) ex
A:Reference number: S68679; MID:96197801
A:Accession: S68679
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-365 <STA>
A:Cross-references: EMBL:X96597; NID:g1296631; PIDN:CA65415.1; PID:e229605; PID:g129
C:Superfamily: ATP receptor P2u
C:Keywords: G protein-coupled receptor

Query Match 99.7%; Score 1940; DB 2; Length 365;
Best Local Similarity 100.0%; Pred. No. 3.8e-164;
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASTESSLLRLSLGLSPGSSFEVLDQWDEDFKFLIPSYAVVYVGLGNAPTLMF 60
DB 1 MASTESSLLRLSLGLSPGSSFEVLDQWDEDFKFLIPSYAVVYVGLGNAPTLMF 60
QY 61 IFRLRPMDATATYMFHLALSDTLVLSPLTIYYAAHNMPPGTICKRFRLFYWNL 120
DB 61 IFRLRPMDATATYMFHLALSDTLVLSPLTIYYAAHNMPPGTICKRFRLFYWNL 120
QY 121 CSYLFETCISVHRHYGICHPRLALRMGRPRLAGLCLAVLVVAGCLVPLFFVTSNKG 180
DB 121 CSYLFETCISVHRHYGICHPRLALRMGRPRLAGLCLAVLVVAGCLVPLFFVTSNKG 180
QY 181 TTYVCHDTRPEEFDFHYHSSAVMGLFGVPCLVTVYCGLMARLYOPLPSSAOSSSR 240
DB 181 TTYVCHDTRPEEFDFHYHSSAVMGLFGVPCLVTVYCGLMARLYOPLPSSAOSSSR 240
QY 241 LRSRLTAVLVFAVCFVPHRTTITTYLRLLEADCVLNVVNVVYVTRPLAANSC 300
DB 241 LRSRLTAVLVFAVCFVPHRTTITTYLRLLEADCVLNVVNVVYVTRPLAANSC 300
QY 301 LDFVLYLTGDKYRROLRLCGGKQPRTAASLALVSLPDSRCRMATPDSSCSTP 360
DB 301 LDFVLYLTGDKYRROLRLCGGKQPRTAASLALVSLPDSRCRMATPDSSCSTP 360
QY 361 RADR 364
DB 361 RADR 364
RESULT 2
A47556
ATP receptor P2u - mouse
C:Species: Mus musculus (house mouse)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999
C:Accession: A47556

R:Lustig, K.D.; Shiau, A.K.; Brake, A.J.; Julius, D.

Proc. Natl. Acad. Sci. U.S.A. 90, 5113-5117, 1993

A:Title: Expression cloning of an ATP receptor from mouse neuroblastoma cells.

A:Reference number: A47556; MID:93281707

A:Accession: A47556

A:Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-373 <US>

A:Cross-references: GB:114751; NID:9309457; PIDN:AAA9871.1; PID:9309458

A:Superfamily: ATP receptor P2u

C:Keywords: transmembrane protein

Query Match

Best Local Similarity 49.5%; Score 962.5; DB 2; Length 373;
Pred. No. 9,7e-78;

Matches 188; Conservative 46; Mismatches 97; Indels 19; Gaps 3;

22 EVELDCWFDEDFKFLIPVSAVAVFVLGLNAPTLMLFIRLPRNATATYMHLLSD 81

20 ELGKCRNFEDFKVLLPVSYGVVGLCLINVALTIFLCRLKTMNASTYMHLLAVSD 79

82 TLVYLSLPLTLYYAAHNMHPFETICKFVRFLEFYNLYCSVLEFLTCISVHRIGICHL 141

80 SLVAASLPLLYYYARGDHPFSTVLCKLVRFLEYTLNLYCSIFLTCISVHRIGICHL 139

142 RALRMGRPRLAGLCLAVLWLVAGCLVPLNFVYTTSNKGTTLVCHDTTREPFDHYHFS 201

140 HSLRMGRARARVAVVWLVLAQAPVLYFVYTTSGRTITCHDSARELFSEFVAYS 199

202 SAVMGLFVPCVLTIVCYGMARLQPLPGSNOSSRL--RSLRTIYAVLVFANCFV 259

200 SVMLGLFVAVPVLIVCYLMARLKPAYGTGGLPRKRRKSVRTIALVLAVALCF 259

260 PFHTRTIYVLAIRLEADCVLNIVNVYKTRPLASANSCLDPVLYLTGDKYRQLRQ 319

260 PFHTRTIYVLSFSLDLSCHTLNAINAKYTRPLASANSCLDPVLYLTGDKYRQLRQ 319

320 LCGGGRKQP-----RTAASSLALVSPEDSSCRMAATPQDS 355

320 AKPTEPTPSQARRKLGRLHPRNTRKDL---SVSSDDSRRESTRPAGS 366

RESULT 3

A54946

P-2U nucleotide receptor - human

C:Species: Homo sapiens (man)

C:Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 17-Mar-1999

C:Accession: A54946

R:Part: C.E.; Sullivan, D.M.; Paradiso, A.M.; Lazarowski, E.R.; Burch, L.H.; Olsen, J.C.

Proc. Natl. Acad. Sci. U.S.A. 91, 3275-3279, 1994

A:Title: Cloning and expression of a human P-2U nucleotide receptor, a target for cystic

A:Reference number: A54946; MID:94211846

A:Accession: A54946

A:Status: preliminary

A:Molecule type: mRNA, protein

A:Residues: 1-375 <PAR>

A:Cross-references: GB:U07225

A:Note: parts of this sequence were confirmed by protein sequencing

C:Genetics:

A:Gene: GDB:P2RY2; HP2U; P2U

A:Cross-references: GDB:362713; OMIM:600041

A:Map position: 11q13.5-11q14.1

C:Superfamily: ATP receptor P2u

C:Keywords: G, protein-coupled receptor; transmembrane protein

Query Match 48.0%; Score 934; DB 2; Length 375;

Best Local Similarity 58.5%; Pred. No. 3.2e-75;

Matches 183; Conservative 41; Mismatches 85; Indels 4; Gaps 3;

22 EVELDCWFDEDFKFLIPVSAVAVFVLGLNAPTLMLFIRLPRNATATYMHLLSD 81

20 ELGKCRNFEDFKVLLPVSYGVVGLCLINVALTIFLCRLKTMNASTYMHLLAVSD 79

QY 82 TLVYLSLPLTLYYAAHNMHPFETICKFVRFLEFYNLYCSVLEFLTCISVHRIGICHL 141

80 ALVAASLPLLYYYARGDHPFSTVLCKLVRFLEYTLNLYCSIFLTCISVHRIGICHL 139

QY 142 RALRMGRPRLAGLCLAVLWLVAGCLVPLNFVYTTSNKGTTLVCHDTTREPFDHYHFS 201

140 HSLRMGRARARVAVVWLVLAQAPVLYFVYTTSGRTITCHDSARELFSEFVAYS 198

QY 202 SAVMGLFVPCVLTIVCYGMARLQPLPGSNOSSRL--RSLRTIYAVLVFANCFV 259

199 SVMLGLFVAVPVLIVCYLMARLKPAYGTGGLPRKRRKSVRTIALVLAVALCF 258

QY 260 PFHTRTIYVLAIRLEADCVLNIVNVYKTRPLASANSCLDPVLYLTGDKYRQLRQ 319

259 PFHTRTIYVLSFSLDLSCHTLNAINAKYTRPLASANSCLDPVLYLTGDKYRQLRQ 317

QY 320 LCGGGRKQPRTAA 332

318 AKPTEPTPSQARRKLGRLHPRNTRKDL---SVSSDDSRRESTRPAGS 330

Db

Db

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

QY

```

RESULT      5
S54253
purinergic receptor - human
C:Species: Homo sapiens (man)
C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 24-Sep-1999
C:Accession: S54253
R:Leon, C.; Vial, C.; Cazenave, J.; Gachet, C.
submitted to the EMBL data Library, May 1995.
A:Description: Cloning of a human putative P2Y receptor.
A:Reference number: S54253
A:Accession: S54253
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1372 <LEO>
A:Cross-references: EMBL:Z49205; NID:g798835; PIDN:CAA83066.1; PID:g798836
C:Genetics:
A:Gene: GDB:P2RY1
A:Cross-references: GDB:677125; OMIM:601167
A:Map position: 3pter-3qter
C:Superfamily: ATP receptor P2u

```

```

QY 1 / GPSS-----EVELDCWFE--DEFILLPVSAYVEVLGALNAPMTMLIFR 63
D 20 GPGSSMGSTVAATAAASSSEKCALTKGFEYLLPAVILVFIIGLGSVALMVEFH 79
QY 64 LRPMDATNYFHLATSPILVLSLPLILYYAAHNHMPSTELCKVRELEFYNLYCSV 123
D 80 MKPMSGISVYFNALADPLVLLPALIFEFENKTKWIDIEADACKQIRLEFHNLY-GI 138
QY 124 LELTCISVHRYLIGICHPRLALRMGRPLRGLLCA--VMEVAGCLVPNLFYVTS-NKG 180
D 139 LELTCISVHRYSGVVYPLKSL--GRLKKNAKICISVYMLVVAISPLFYSTGYRKN 196
QY 181 TTVCHDTTRREEDHYVHSESAVMGLFGVPCLVLYCGIIMARRL-YOPLPGSAOSS 238
D 197 KTICTDITSDERISRFIYSMCTTVAFCVPLVILIGCGILVIRALITYRDL--NSPL 253
QY 240 RLRSRLTAIVLVFAVCEPFIHTRIYYLARI--LEADCRLVNTVNVYKTRPLAS 296
D 254 RRKSIYLVILVFAVSYIPFHYMKTMNLBARDEQTPAMCANDRYVNTQYTRBLAS 313
QY 297 ANSGLDPLVLLTGDKYRQL---RQLCGGGRQKQPTAASSLAVSLVPE 342
D 314 LNSCVDPILVFLAGDTERRRRLSRATRRASRSEANLQSKEDMTLNLPE 363

```

```

Query Match          32.3%, Score 628; DB 2; Length 362;
Best Local Similarity 39.4%, Pred. No. 3.5e-48;
Matches 137; Conservative 62; Mismatches 133; Indels 16; Gaps
7,

QY      4 TESSLLRSLGLSGPGSGSEVELDCMFEDDFKRLPVSAAVVFVGLGNAETPLWLFIR 63
Db      13 TQPELLAG-GWMAAGNATKCSLT---KGFQGYVLPPTYLLVFIITGLGNSAIAIMWFVH 68
QY      64 LRPWNAATVMEHNLSTLYLVLSLPTLIYYAAAHNHPEFGTEICKEVRFELVMNLKCV 123
Db      69 MRPMGSGIVYMENMLAOLFVYLTLPALIFEYFNKTDWIFGVGMCKLRFETIHVNLGSI 128
QY      124 LFLATCSVHRYLGICHPRLALRGKRPRLAGLLCLLAIVYVAGCLVNPQFEVITS-NKGT 182
Db      129 LFLNCSIVHRTGVVHPLKSLGRKKRNAAVYSSVMAVLVAIVAPILFYSGTVRRNKT 188
QY      183 VLCHDTTPREEFDHYVHSESAVMGLLFVGPCLVTLVYCYGLMARRL-YQPLPGSAOSSRL 241
Db      189 ITCYDTTADETLRSFYFVSMCTVMEFCIFVLIGCYGLVIALIYDLD--NSPLRR 245
QY      242 RSLRLIAVLVLEAFNCFPFHITRIYYLARL---LEADCRVLNVNVVYKYTRPLASAN 298
Db      246 KSIYLVIVLVLEAFVASYLPEFVMKTLNLRARLDQTPOMCAFENDCVAYVAYQVTRGLASIN 305
QY      299 SCLDPVLLITLGTGKKYRQL---ROLGCGGKPPQPTFAASSLALVSLPE 342
Db      306 SCVDPILXELAGDTRFRRLSRATKRSSRRSEPNQSKSEEMTLNLTLE 353

```

RESULT 7
JC4162
P2Y receptor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 12-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 24-Sep-1999
C:Accession: JC4162
R:Henderson, D.U.; Elliott, D.G.; Smith, G.M.; Webb, T.E.; Dainty, I.A.
Biochem. Biophys. Res. Commun. 212, 648-656, 1995
A>Title: Cloning and characterisation of a bovine P2Y receptor.
A:Reference number: JC4162; MUID:95352058
A:Accession: JC4162
A:Molecule type: mRNA
A:Residues: 1-373 <HEN>
A:Cross-references: EMBL:X87628; NID:g1032484; PDB:CAA60958.1; PID:g1032484
A:Experimental source: aortic endothelial cell
C:Genetics:
A:Gene: bopy2y
C:Superfamily: ATP receptor P2u
C:Keywords: glycoprotein; phosphoprotein; receptor; transmembrane protein
F:5-77/Domain: transmembrane #status predicted <TM1>
F:88-111/Domain: transmembrane #status predicted <TM2>
F:114-150/Domain: transmembrane #status predicted <TM3>
F:111-191/Domain: transmembrane #status predicted <TM4>
F:214-237/Domain: transmembrane #status predicted <TM5>
F:261-282/Domain: transmembrane #status predicted <TM6>
F:305-328/Domain: transmembrane #status predicted <TM7>

F:193-216/Domain: transmembrane #status predicted <TM6>
F:241-264/Domain: transmembrane #status predicted <TM6>
F:283-305/Domain: transmembrane #status predicted <TM7>
F:5,173/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 30.1%; Score 586; DB 2; Length 328;
Best Local Similarity 42.5%; Pred. No. 1,6e-44;
Matches 135; Conservative 37; Mismatches 128; Indels 18; Gaps 6;

QY 10 RSIGLSPGSGSEVELDCEWDEDFKILLPVSAVAVVGLGAINAPTLMLFIRLRPMWA 69
DB 9 QALGLP-----TTCYREHFKOLLPPYSAVLAAGLPINICVITQICTSRRLNR 60
QY 70 TATYVHIALDSTLYVLSPTLYYYAAHNNHMPGTEICKFVRFYVNLGCVLELTCT 129
DB 61 TAVYTLNALADLADLVACSLPLITLYNAAGDHMFEGDFACGLVAFLEFVLANHGISLLELTCT 120
QY 130 SVHRYLGICHPRLALW---GPRRLAGLCLAVLVVAGCLVNLFFVTSNKGTVLCH 186
DB 121 SFQRYLGICHPRLAP--WHRGGRRAAMLVCAVAVLAVTTQCLTAIFATIGIQNRFTVCY 178
QY 187 DTRPEEDFHYVHSSAVVAGLFGVPCVLTLYCYGLMARLYQ---PLPGSAOSSRLRS 243
DB 179 DLSPPALAHYHYPYGMALVYIGFLFPALALACVCLLACRLCRQDGPAPPAQ-ERRGR 237
QY 244 LPTAVVLTAVFVAVCEVPHITRTIYLLARLE-ADCRVLINVVVYKTRPLASANSCLD 302
DB 238 ANNAVYVAAPAFISFLPHITKATVAVSTPGVCTVLEAFPAAKGTRPFASANSVLD 237
QY 303 PVLVLTGDKYRRQLKQL 320
DB 298 PILFTYTKRFRRRPHEL 315

RESULT 11

150241

G:protein-coupled receptor 6H1 - chicken
N:Alternate names: purinoceptor 6H1
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: I50241; J04618
R:Kaplan, M.H.; Smith, D.I.; Sundick, R.S.
J:Immunol. 151, 628-636, 1993
A:Title: Identification of a G protein coupled receptor induced in activated T cells.
A:Reference number: I50241; MUID:93329058
A:Accession: I50241
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-308 <KAP>
A:Cross-references: GB:I06109; NID:9304383; PIDN:AA06587.1; PID:9304384
R:Webb, T.E.; Kaplan, M.G.; Barnard, E.A.
Biochem. Biophys. Res. Commun. 219, 105-110, 1996
A:Title: Identification of 6H1 as a P2Y purinoceptor: P2Y5.
A:Reference number: J04618; MUID:96190677
A:Accession: J04618
A:Molecule type: mRNA
A:Residues: 1-308 <WEB>
A:Cross-references: DDBJ:D49712
A:Experimental source: T-cells
C:Comment: This receptor plays a role in T-cell activation.
C:Genetics:
A:Gene: P2Y5
A:Superfamily: ATP receptor P2U
C:Keywords: G protein-coupled receptor; transmembrane protein
F:15-40/Domain: transmembrane #status predicted <TM1>
F:51-74/Domain: transmembrane #status predicted <TM2>
F:89-109/Domain: transmembrane #status predicted <TM4>
F:133-153/Domain: transmembrane #status predicted <TM5>
F:177-201/Domain: transmembrane #status predicted <TM5>
F:227-248/Domain: transmembrane #status predicted <TM6>
F:269-292/Domain: transmembrane #status predicted <TM7>

Query Match 25.8%; Score 501.5; DB 2; Length 308;
Best Local Similarity 36.9%; Pred. No. 4.5e-37;
Matches 108; Conservative 56; Mismatches 124; Indels 5; Gaps 4;

QY 23 VEIDCWDFDEDFKILLPVSAVAVVGLGAINAPTLMLFIRLRPMWATATYVHIALSDT 82
DB 2 VSSNCSDEDSFKYTLGCVFSAVFLGLIANCAIYIFETTLKVRNETTYMLNLSIDL 61
QY 83 LVYLSLPTLYYYAAHNNHMPGTEICKFVRFYVNLGCVLELTCTSVHRYLGICHPRL 142
DB 62 LEVFTLPFRIFYVARN-WFEGDVLCKISVTLFTYTNMGISLFLTCISVDRFLAIHPFR 120
QY 143 ALRMGRPRLAGLCLAVLVVAGCLVNLFFVTSNKGTV--LCGHTTRPEEDHYVH 200
DB 121 SKLRKRNRARIYCAVAVITVLAGSTPASFOSTNRONTQRTCEBNPESTWKTLYLR 180
QY 201 SSAVMGLL-FGVPCVLTLYCYGLMARLYQPLPGSAOSSRLSLRTIAVLTAVFVAVCEV 259
DB 181 IVYFIEIVGFIFLILNVCTWVLRFLNKLPLSLRNKLSKKVLMIFVHLVIFCCFV 240
QY 260 PHITRTIYLLARLE-ADCRVLINVVYKTRPLASANSCLDPLVLYLTGD 311
DB 241 PYNITLLIXLMRTQWNCVSTAVRTMYPVTLCAVSNCCPDIYYFTSD 293

RESULT 12

T09508

Intron 17 purinergic receptor P2Y5 - human
N:Alternate names: G-protein coupled receptor
C:Species: Homo sapiens (man)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T09508
R:Bohm, S.K.; Trumpp, A.; Khitin, L.M.; Kong, W.; Payan, D.G.; Bunnett, N.W.
submitted to the EMBL Data Library, April 1997
A:Description: The human purinergic receptor P2Y5 is encoded in Intron 17 of the ret1
A:Reference number: Z16705
A:Accession: T09508
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-344 <BOH>
A:Cross-references: EMBL:AF000546; NID:92232068; PID:92232069
C:Genetics:
A:Map position: 13
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 24.6%; Score 479.5; DB 2; Length 344;
Best Local Similarity 33.4%; Pred. No. 4.4e-35;
Matches 99; Conservative 66; Mismatches 126; Indels 5; Gaps 4;

QY 27 CWFDEDFKILLPVSAVAVVGLGAINAPTLMLFIRLRPMWATATYVHIALSDTLYLV 86
DB 9 CFYNDSEFKYTLGCVFSAVFLGLIANCAIYIFICVLKVRNETTYMLNLSIDLFFV 68
QY 87 SLPTLYYYAAHNNHMPGTEICKFVRFYVNLGCVLELTCTSVHRYLGICHPRLALRM 146
DB 69 TLPRIFYFTTRN-WFEGDVLCKISVTLFTYTNMGISLFLTCISVDRFLAIHPFRSKUL 127
QY 147 GRPRLAGLCLAVLVVAGCLVNLFFVTSNKGTV--LCGHTTRPEEDHYVHSSAV 204
DB 128 RTKRNKAIYCTGWLTVIGSAPAVVOSTHSGNNASEACEFNPEDATWKTLYLSRIYF 187
QY 205 MGLL-FGVPCVLTLYCYGLMARLYQPLPGSAOSSRLSLRTIAVLTAVFVAVCEVPHI 263
DB 188 IELVGFIFILINVTOSMWLKLTRPVLLSRKIKTKVLMKIFPHLLIFCCFVPHYI 247
QY 264 TRITTYLLARLE-ADCRVLINVVYKTRPLASANSCLDPLVLYLTGDKYRRQLR 318
DB 248 NLIVSLVTRQTFVNCVAAVATWYPTILCAVSNCCPDIYYFTSDIYNSIK 303

RESULT 13

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 29, 1999, 19:43:41 ; Search time 26.5 seconds
(without alignments)
400.618 Million cell updates/sec

Title: US-09-077-173a-2

Perfect score: 1946
Sequence: 1 NASTESSLRSLRSLSPGPGS.....CWMATPDSSCSTPEADRF 365

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database : SwissProt_38.*

Word size : 0

Number of hits that pass the threshold : 80000

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1940	99.7	365	1	P2Y4_HUMAN	P51382 homo sapien
2	1007.5	51.8	337	1	P2Y8_XENLA	P79928 xenopus lae
3	965	49.6	377	1	P2YR_HUMAN	P41331 homo sapien
4	962.5	49.5	373	1	P2UR_MOUSE	P35383 mus musculu
5	950	48.8	373	1	P2UR_RAT	P41332 rattus norv
6	641.5	33.0	373	1	P2YR_HUMAN	P47900 homo sapien
7	631	32.4	362	1	P2YR_MELGA	P49652 meleagris g
8	628	32.3	362	1	P2YR_CHICK	P34996 gallus gall
9	621	31.9	373	1	P2YR_BOVIN	P48042 bos taurus
10	616	31.7	373	1	P2YR_RAT	P49651 rattus norv
11	614	31.6	373	1	P2YR_MOUSE	P49650 mus musculu
12	603	31.0	328	1	P2Y3_CHICK	Q98907 gallus gall
13	588	30.2	328	1	P2Y6_RAT	Q63371 rattus norv
14	586	30.1	328	1	P2Y6_HUMAN	Q15077 homo sapien
15	510	26.2	339	1	GPRH_HUMAN	Q13304 homo sapien
16	501.5	25.8	308	1	P2Y5_CHICK	P32250 gallus gall
17	479.5	24.6	344	1	P2Y5_HUMAN	P43657 homo sapien
18	449.5	23.1	370	1	P2Y9_HUMAN	Q99677 homo sapien
19	426	21.9	397	1	PAR2_RAT	O63645 rattus norv
20	423	21.7	399	1	PAR2_MOUSE	P55086 mus musculu
21	411.5	21.1	420	1	THRR_XENLA	P47749 xenopus lae
22	403	20.7	397	1	THRR_HUMAN	P55085 homo sapien
23	397.5	20.4	425	1	THRR_HUMAN	P25116 homo sapien
24	393.5	20.2	425	1	THRR_PAPHA	P56488 papio hamad
25	393	20.2	430	1	THRR_MOUSE	P30558 mus musculu
26	388	19.9	432	1	THRR_RAT	P26824 rattus norv
27	387.5	19.9	428	1	THRR_CRILLO	Q00991 cricetus
28	385	19.8	372	1	OPRD_RAT	P33333 rattus norv
29	384	19.7	372	1	OPRD_HUMAN	P41143 homo sapien
30	383	19.7	361	1	EB12_HUMAN	P32249 homo sapien
31	382.5	19.6	401	1	AG2R_XENLA	P32303 xenopus lae
32	380.5	19.5	401	1	OPRM_PIG	Q95247 sus scrofa
33	380	19.5	401	1	OPRM_BOVIN	P79350 bos taurus
34	378.5	19.5	400	1	OPRM_HUMAN	P35372 homo sapien
35	377	19.4	372	1	OPRD_MOUSE	P32300 mus musculu
36	376.5	19.3	398	1	OPRM_RAT	P33335 rattus norv
37	375.5	19.3	374	1	PAR3_HUMAN	Q00254 homo sapien
38	375	19.3	368	1	CCR3_HUMAN	P49682 homo sapien
39	371.5	19.1	398	1	OPRM_MOUSE	P42866 mus musculu

40	371.5	19.1	342	1	PAPR_HUMAN	P25105 homo sapien
41	369.5	19.0	359	1	AG2R_MOUSE	P29754 mus musculu
42	367	18.9	357	1	GC96_HUMAN	P51686 homo sapien
43	365	18.8	363	1	SSRS_HUMAN	P35346 homo sapien
44	364.5	18.7	359	1	AG2R_RAT	P25095 rattus norv
45	363	18.7	359	1	AG2S_RAT	P29089 rattus norv

ALIGNMENTS

```
RESULT 1
P2Y4_HUMAN
ID P2Y4_HUMAN STANDARD: PRT; 365 AA.
AC P51382;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE URIDINE NUCLEOTIDE RECEPTOR (P2Y) (P2Y4).
GN P2RY4 OR NRU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96125055.
RA COMMENT D., PIROTON S., PARMENTIER M., BOEYNAENS J.-M.;
RT "Cloning and functional expression of a human uridine nucleotide
RT receptor."
RL J. Biol. Chem. 270:30849-30852(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96125054.
RA NGUYEN T., ERB L., WEISMAN G.A., MARCHESE A., HENG H.H.Q.,
RA GARRAD R.C., GEORGE S.R., TURNER J.T., O'DONN B.F.;
RT "Cloning, expression, and chromosomal localization of the human
RT uridine nucleotide receptor gene."
RL J. Biol. Chem. 270:30845-30848(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE-PANCREAS;
RX MEDLINE: 96197801.
RA STEAM N.J., KLOMP J., VAN DER HEUVEL M., OLIVE W.;
RT "Molecular cloning and characterization of a novel orphan receptor
RT (P2P) expressed in human pancreas that shows high structural homology
RT to the P2U purinoceptor."
RL FEBS Lett. 384:260-264(1996).
CC -!- FUNCTION: RECEPTOR FOR UTP AND UDP, NOT ACTIVATED BY ATP OR ADP.
CC SEEMS TO MEDIATE ITS ACTION VIA ACTIVATION OF A
CC PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: PANCREAS.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
CC EMBL: X91852; CAA62963.1; -
CC EMBL: U40223; AAC50347.1; -
CC EMBL: X96597; CAA65415.1; -
CC HSSP: P34996; 1DDP.
CC GCRDB; GCR_1952; -
CC GCRDB; GCR_2065; -
CC GCRDB; GCR_2085; -
CC MIM: 300038; -
CC PFMW: PF00001; 7tm_1; 1
CC PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
CC DR
CC DR G-protein coupled receptor; Transmembrane; Glycoprotein.
CC KW
```

```

FT DOMAIN 1 34 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 35 61 1 (POTENTIAL).
FT DOMAIN 62 72 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 73 95 2 (POTENTIAL).
FT DOMAIN 96 112 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 113 131 3 (POTENTIAL).
FT DOMAIN 132 154 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 155 174 4 (POTENTIAL).
FT DOMAIN 175 196 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 197 222 5 (POTENTIAL).
FT DOMAIN 223 246 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 247 269 6 (POTENTIAL).
FT DOMAIN 270 287 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 288 309 7 (POTENTIAL).
FT DOMAIN 310 365 CYTOPLASMIC (POTENTIAL).
FT DISULFID 108 185 BY SIMILARITY.
FT CONFLICT 86 86 L -> V (IN REF. 2).
FT CONFLICT 178 178 N -> T (IN REF. 2).
FT CONFLICT 234 234 S -> A (IN REF. 2).
SQ SEQUENCE 365 AA; 40963 MW; ABE553A7 CRC32;

```

```

Query Match 99.7%; Score 1940; DB 1; Length 365;
Best Local Similarity 100.0%; Pred. No. 2.5e-129;
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MASTESSLLSLGSPGSSSEVELDCWFEDEFFKILLPVSAVAVVGLGLNAPLTLMLF 60
DB 1 MASTESSLLSLGSPGSSSEVELDCWFEDEFFKILLPVSAVAVVGLGLNAPLTLMLF 60
QY 61 IFLRLPMDATATYMFHALSDTLVLSPLTIYYAANMPETETCKVRFLEYNNLY 120
DB 61 IFLRLPMDATATYMFHALSDTLVLSPLTIYYAANMPETETCKVRFLEYNNLY 120
QY 121 CAVLFTLCISVHRYLGICHPRLALRMGRPRLAGLCLAVLVVAGCLVPLFEVTTSNKG 180
DB 121 CAVLFTLCISVHRYLGICHPRLALRMGRPRLAGLCLAVLVVAGCLVPLFEVTTSNKG 180
QY 181 TTYVLCDDTTPREEDHRYHSSAVMGLLFVPCVLTLYVCGLMARLYQPLPGSAGSSSR 240
DB 181 TTYVLCDDTTPREEDHRYHSSAVMGLLFVPCVLTLYVCGLMARLYQPLPGSAGSSSR 240
QY 241 LRLSLRTIAVAVLFAVCEVFPHITRTIYARLLEADCRVLNTVNVYKTRPLASANSNC 300
DB 241 LRLSLRTIAVAVLFAVCEVFPHITRTIYARLLEADCRVLNTVNVYKTRPLASANSNC 300
QY 301 LDPVLYLLTGDKYRQLRQLCGGGKPPQRTAASSIALVSLPDSGCRWATPODSSGCTP 360
DB 301 LDPVLYLLTGDKYRQLRQLCGGGKPPQRTAASSIALVSLPDSGCRWATPODSSGCTP 360
QY 361 RADR 364
DB 361 RADR 364

```

```

RESULT 2
P2Y8_XENLA STANDARD; PRT; 537 AA.
AC P79928;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE P2Y PURINOCEPTOR 8 (P2Y8).
GN P2Y8.
OS Xenopus laevis (African clawed frog).
OC Batrachia; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
SEQUENCE FROM N.A.
RC TISSUE-NEURAL PLATE;
RX MEDLINE: 97284734.
RA BOGDANOV Y.D.; DALE L.; KING B.F.; WHITLOCK N.; BURNSTOCK G.;

```

```

RT "Early expression of a novel nucleotide receptor in the neural plate
of Xenopus embryos."
RL J. Biol. Chem. 272:12583-12590(1997).
CC -1- FUNCTION: RECEPTOR FOR EXTRACELLULAR ATP, UTP, CTP, GTP AND ITP.
CC THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEIN WHICH
CC ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC MAY PLAY A KEY ROLE IN THE EARLY DEVELOPMENT OF NEURAL TISSUE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: X89953; CAA68213.1; -
DR HSSP: P34996; 1DDO.
DR GCRDB: GCR_1419; -
DR PFAM: PF00001; 7tm1; 1.
DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 49 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 50 70 1 (POTENTIAL).
FT DOMAIN 71 79 2 (POTENTIAL).
FT TRANSSEM 80 100 2 (POTENTIAL).
FT DOMAIN 101 118 3 (POTENTIAL).
FT TRANSSEM 119 139 3 (POTENTIAL).
FT DOMAIN 140 161 4 (POTENTIAL).
FT TRANSSEM 162 182 4 (POTENTIAL).
FT DOMAIN 183 210 5 (POTENTIAL).
FT TRANSSEM 211 231 5 (POTENTIAL).
FT DOMAIN 232 254 6 (POTENTIAL).
FT TRANSSEM 255 275 6 (POTENTIAL).
FT DOMAIN 276 276 7 (POTENTIAL).
FT TRANSSEM 276 276 7 (POTENTIAL).
FT DISULFID 116 193 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 26 26 BY SIMILARITY.
FT CARBOHYD 29 29 POTENTIAL.
SQ SEQUENCE 537 AA; 62024 MW; DB5A8743 CRC32;

```

```

Query Match 51.8%; Score 1007.5; DB 1; Length 537;
Best Local Similarity 55.9%; Pred. No. 6.8e-64;
Matches 190; Conservative 50; Mismatches 83; Indels 17; Gaps 4;

```

```

QY 20 SSEVELDCWFEDEFFKILLPVSAVAVVGLGLNAPLTLMLFIRLPMDATATYMFHAL 79
DB 28 TNDTEDIQVDFGDFKILLPVSAVAVVGLGLNAPLTLMLFIRLPMDATATYMFHAL 87
QY 80 SPTLYVLSPLTIYYAANMPETETCKVRFLEYNNLYCISVHRYLGICH 139
DB 88 SPTLYVLSPLTIYYAANMPETETCKVRFLEYNNLYCISVHRYLGICH 147
QY 140 PLRLRMGRPRLAGLCLAVLVVAGCLVPLFEVTTSNKGTTVLCDDTTPREEDHYH 199
DB 148 PLRLRMGRPRLAGLCLAVLVVAGCLVPLFEVTTSNKGTTVLCDDTTPREEDHYH 207
QY 200 FSSAVAGLFGVPCVLTLYVCGLMARLYQPLPGSAGSSSRSLRTIAVAVLTVFAV 256
DB 208 YSTALNCLLFGPCILINAGCYGLMTRLEKPLVSGNQTLPYKRSIKTITFVIAFAI 267
QY 257 CEPFHITRTIYARLLEADCRVLNTVNVYKTRPLASANSCLDPVLYLLTGKYSRQ 316
DB 268 CEPFHITRTIYARLLEADCRVLNTVNVYKTRPLASANSCLDPVLYLLTGKYSRQ 327
QY 317 L-----ROLGGGKFO--PRTAASSIALVSLPE 342
DB 328 LIRTVRRSSVYNRCHMTNHPOTEPHMTAGPLVISAEE 367

```

RESULT	3			
P2UR_HUMAN				
ID	P2UR_HUMAN	STANDARD:	PRT:	377 AA.
AC	P41231.			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	P2UR PURINERGIC RECEPTOR 1 (ATP RECEPTOR) (P2U1) (PURINERGIC RECEPTOR).			
GN	P2UR2 OR P2UR1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-AIRWAY EPITHELIOU;			
RX	MEDLINE; 94211846.			
RA	PARC C.E., SULLIVAN D.M., PARADISO A.M., LAZAROWSKI E.R., BURCH L.H.,			
RA	OLSEN J.C., ERB L., WEISMAN G.A., BOUCHER R.C., TURNER J.T.;			
RT	"Cloning and expression of a human P2U nucleotide receptor, a target			
RT	for cystic fibrosis pharmacotherapy."			
RL	Proc. Natl. Acad. Sci. U.S.A. 91:3275-3279(1994).			
CC	-1- FUNCTION: RECEPTOR FOR EXTRACELLULAR UTP = ATP > ATP-GAMMA-S >> 2-			
CC	METHYLTHIO-ATP = ADP. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY			
CC	G PROTEINS WHICH ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND			
CC	MESSNGER SYSTEM.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-1- TISSUE SPECIFICITY: SPLEEN, TESTIS, KIDNEY, LIVER, LUNG, HEART AND			
CC	BRAIN.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-----			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION			
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -			
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS			
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY			
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL			
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/			
CC	OR SEND AN EMAIL TO license@isb-sib.ch).			
CC	-----			
DR	EMBL; U07225; AAC04923.1; -.			
DR	EMBL; S74902; AAB32966.1; -.			
DR	PIR; A54946; A54946.			
DR	HSSP; P34996; 1DD.			
DR	GCRDB; GCR 1008; -.			
DR	MIM; 600041; -.			
DR	PFAM; PF00001; 7tm_1; 1.			
KW	PROSITE; PS00237; G-PROTEIN RECEPTOR; 1.			
KW	G-protein coupled receptor; Transmembrane; Glycoprotein;			
FT	phosphorylation.			
FT	DOMAIN 1			
FT	TRANSMEM 33			
FT	TRANSMEM 59			
FT	TRANSMEM 60			
FT	TRANSMEM 71			
FT	TRANSMEM 70			
FT	TRANSMEM 93			
FT	DOMAIN 94			
FT	TRANSMEM 110			
FT	TRANSMEM 111			
FT	TRANSMEM 129			
FT	TRANSMEM 130			
FT	TRANSMEM 152			
FT	TRANSMEM 153			
FT	TRANSMEM 172			
FT	TRANSMEM 173			
FT	TRANSMEM 194			
FT	TRANSMEM 195			
FT	TRANSMEM 220			
FT	TRANSMEM 221			
FT	TRANSMEM 246			
FT	TRANSMEM 247			
FT	TRANSMEM 269			
FT	TRANSMEM 270			
FT	TRANSMEM 287			
FT	TRANSMEM 288			
FT	TRANSMEM 309			
FT	DOMAIN 310			
FT	DOMAIN 377			
FT	CARBOHYD 9			
FT	9			
FT	POTENTIAL.			
FT	EXTRACELLULAR (POTENTIAL).			
FT	1 (POTENTIAL).			
FT	CYTOPLASMIC (POTENTIAL).			
FT	2 (POTENTIAL).			
FT	EXTRACELLULAR (POTENTIAL).			
FT	3 (POTENTIAL).			
FT	CYTOPLASMIC (POTENTIAL).			
FT	4 (POTENTIAL).			
FT	EXTRACELLULAR (POTENTIAL).			
FT	5 (POTENTIAL).			
FT	CYTOPLASMIC (POTENTIAL).			
FT	6 (POTENTIAL).			
FT	EXTRACELLULAR (POTENTIAL).			
FT	7 (POTENTIAL).			
FT	CYTOPLASMIC (POTENTIAL).			
FT	POTENTIAL.			

FT	CARBOHYD	-	13	POTENTIAL.
FT	DISULFID	106	183	BY SIMILARITY.
SEQ	SEQUENCE	377 AA.	42277 MM.	4C9ACT77C CRC32;
Query Match		49.6%;	Score 965;	DB 1; Length 377;
Best Local Similarity		59.1%;	Pred. No. 4,7e-61;	
Matches 185;	Conservative	40;	Mismatches 86;	Indels 2; Gaps 1
QY	22	EVELDCEWDEDEKFIILLPVSYAVVFLVGLGNAPLMLFIFRLRPWDATATYMFHLALSD	81	
DB	20	ELGYRCRFENEDKRYLLLPVSYGVCVGLGLNAVALYIRLCKTKTNASTYMFHLAVSD	79	
QY	82	TLVYLSLPLLYYYAAAHNMHPGTEICKEFRVRLFTWNLKCYSLFELTCISVHRYLGICHP	141	
DB	80	ALYASLPLLYYYVARGDHMPSTYLCIKRLPFLYTNLCISILFELTCISVHRCGLVRLP	139	
QY	142	RALRGPRPLAGLLCLVWLVAAGLVLPNLPVTSNKGTCVLDGDTPEEEDHVAHS	201	
DB	140	RLSLRGRRARYARVAGAVWLVYLAQAPLVYVTTISARQGVYCHDTSAPELFSRFVAYS	199	
QY	202	SAVMGLLEFVPCVLTVLCVGYLMARRLYQPLPGSAOSSRL--RSLRTIVVLTFAVCFV	259	
DB	200	SVMLGLLEAVPFAVLIVCYLMARRLPAGVTSQGLPRAKKSVRTIVAVLAVLACLT	259	
QY	260	PHHTRTIYIYARLLEADCCRVLTIVVYVYKTRPLASANSCLDPVLYLTGSKYRRQLRQ	319	
DB	260	PFHVYRTIYISFRSLDLSCHTLNAINMAYKYTRPLASANSCLDPVLYLAGQRLVRFARD	319	
QY	320	LCGGGKPOPTAA	332	
DB	320	AKPTGSPATPA	332	
RESULT				
P2UR_MOUSE				
ID	P2UR_MOUSE	STANDARD:	PRT:	373 AA.
AC	P35383;	009031;		
DT	01-JUN-1994	(Rel. 29, Created)		
DT	01-JUN-1994	(Rel. 29, Last sequence update)		
DT	01-NOV-1997	(Rel. 35, Last annotation update)		
DE	P2U PURINCEPTOR 1 (ATP RECEPTOR) (P2U1) (PURINERGIC RECEPTOR).			
GN	P2Y2 OR P2UR1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RA	[1]			
RA	SEQUENCE FROM N.A.			
RX	MEDLINE; 93281707.			
RT	LUSTIG K.D., SHAU A.K., BRANE A.J., JULIUS D.;			
RT	"Expression cloning of an ATP receptor from mouse neuroblastoma			
RT	cells.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 90:5113-5117(1993).			
RT	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 96316177.			
RA	ENOMOTO K., FURUYA K., MOORE R.C., YAMAGISHI S., OKA T., MAENO T.;			
RA	"Expression cloning and signal transduction pathway of p2u receptor			
RT	in mammary tumor cells.";			
RT	Biol. Signals 3:9-21(1996).			
CC	-1- FUNCTION: RECEPTOR FOR EXTRACELLULAR UTP - ATP > ATP-GAMMA-S >> 2-			
CC	METHYLTHIO-ATP - ADP. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY			
CC	G PROTEINS WHICH ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND			
CC	MESSENGER SYSTEM.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-1- TISSUE SPECIFICITY: SPLEEN, TESTIS, KIDNEY, LIVER, LUNG, HEART AND			
CC	BRAIN.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

OY 297 ANSCDPVLYLTGDKYRQL-----ROLGGKPPORTNASSLALVSLEPE 342
 DB 315 LNSCDPILYFLAGDTFRRRLSRATRKASRSEANLQSKSEMDTINIILPE 364
 RESULT 7
 P2YR_MELGA STANDARD: PRT: 362 AA.
 AC P49652;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE P2Y PURINORECEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR) (6H1 ORPHAN RECEPTOR).
 GN P2RY1.
 OS Meleagris gallopavo (Common turkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves; Neognathae; Galliformes; Meleagridae; Meleagris.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 94335907.
 RA FILTZ T.M., LI Q., BOYER J.L., NICHOLAS R.A., HARDEN T.K.;
 RT "Expression of a cloned p2y purinergic receptor that couples to
 RL Mol. Pharmacol. 46:8-14(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97382456.
 RA LI Q., SCHACHTER J.B., HARDEN T.K., NICHOLAS R.A.;
 RT "The 6H1 orphan receptor, claimed to be the p2y5 receptor, does not
 RL mediate nucleotide-promoted second messenger responses.";
 RL Biochem. Biophys. Res. Commun. 236:455-460(1997).
 CC -1- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
 CC ATP AND ADP. SEEMS TO MEDIATE ITS ACTION VIA A PERTUSSIS TOXIN
 CC INSENSITIVE G-PROTEIN, PROBABLY BELONGING TO THE GQ FAMILY THAT
 CC ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAINLY FOUND IN BLOOD, BRAIN, AND LUNG. TO A
 CC LESSER EXTENT IN STOMACH, GUT, AND SKELETAL MUSCLE.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U09842; AAA18784.1;
 DR EMBL: AF012103; AAB65428.1;
 DR HSSP: P34996; 1DD.
 DR GCRD: GCR2392;
 DR PRAM: PF00001; 7tm_1; 1.
 DR PROSITE: PS00237; G-PROTEIN RECEPTOR: 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 41
 FT TRANSMEM 63
 FT DOMAIN 64
 FT TRANSMEM 76
 FT TRANSMEM 77
 FT TRANSMEM 98
 FT TRANSMEM 99
 FT TRANSMEM 115
 FT TRANSMEM 116
 FT TRANSMEM 136
 FT TRANSMEM 137
 FT TRANSMEM 155
 FT TRANSMEM 156
 FT TRANSMEM 178
 FT TRANSMEM 208
 FT TRANSMEM 227
 FT TRANSMEM 228
 FT TRANSMEM 254
 FT TRANSMEM 255
 FT TRANSMEM 274
 FT TRANSMEM 275
 FT TRANSMEM 292
 FT TRANSMEM 293
 FT TRANSMEM 318
 FT TRANSMEM 362

FT DISULFID 113 191 BY SIMILARITY.
 FT CARBOHYD 11 11 POTENTIAL.
 FT CARBOHYD 26 26 POTENTIAL.
 FT CARBOHYD 102 102 POTENTIAL.
 FT CARBOHYD 186 186 POTENTIAL.
 SO SEQUENCE 362 AA; 41180 MW; B9D63032 CRC32;
 Query Match 32.4%; Score 631; DB 1; Length 362;
 Best Local Similarity 39.7%; Pred. No. 1,le-37;
 Matches 138; Conservative 61; Mismatches 133; Indels 16; Gaps 7;
 OY 4 TESSLRSLGSPPGSGSEVELDCWDFEDFKILLPVSAVAFVLGGLNAPTLMLTFRR 63
 DB 13 TQPELLAG-GMAAONASTKCLT---KTGFQYFLPYLIVFTTGLGNSVAIWMVEFH 68
 OY 64 LRPDAATYVHFHLSLDTLVLSLPLIYYAANHPFETGECFVRFLEYNLYCSV 123
 DB 69 MRPMGSIYVWFNLADFLVFLPALIFYEKTKDMIGDVMCKLQRFIFHNLGSI 128
 OY 124 LFLTCISVHRTGLCHPLRALRNGRPRLAGLLCLAVLVVAGCLVPMLEFVTS-NGGTT 182
 DB 129 LFLTCISVHRTGVHPLKSLGRLKKNAYVSSLVWALVAVIAPILFYSGTGVRNKT 188
 OY 183 VLCHDTRPEEFHGVHSSAVMGLFEGVPCVLTVOXGLMARL-YQPLPGSSOSSL 241
 DB 189 ITCTDTADELRSFYVSNCTTYVFMCIPIVILGGLVLAIVKLDL--NSPLRR 245
 OY 242 RSLRTIAVLTVAVCPVPHITRTIYIYIARL--LEADCRVLIIVVVKVTRPLASAN 298
 DB 246 KSIYLVIVLTVFAVSLPFPVKKTLRLARLDQFQMKAFNDKYATQVTRGLASLN 305
 OY 299 SCIDPVLYLTGDKYRQL-----ROLGGKPPORTNASSLALVSLEPE 342
 DB 306 SCVDPILYFLAGDTFRRRLSRATRKASRSEANLQSKSEMDTINIILPE 353
 RESULT 8
 P2YR_CHICK STANDARD: PRT: 362 AA.
 ID P2YR_CHICK
 AC P34996;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE P2Y PURINORECEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR).
 GN Gallus gallus (Chicken).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 93285340.
 RA WEBB T.E., SIMON J., KRISHEK B.J., BATESON A.N., SMART T.G.,
 RA KING B.F., BURSTOCK G., BARNARD E.A.;
 RT "Cloning and functional expression of a brain G-protein-coupled ATP
 RT receptor.";
 RL FEBS-Lett. 324:219-225(1993);
 RN [2]
 RP 3D-STRUCTURE MODELING.
 RX MEDLINE: 97026278.
 RA VAN RHEE A.M., FISCHER B., VAN GALEN P.J.M., JACOBSON K.A.;
 RT "Modelling the p2y purinoreceptor using rhodopsin as template.";
 RL Drug Des. Discov. 13:133-140(1995).
 CC -1- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
 CC ATP AND ADP. SEEMS TO MEDIATE ITS ACTION VIA A PERTUSSIS TOXIN
 CC INSENSITIVE G-PROTEIN, PROBABLY BELONGING TO THE GQ FAMILY THAT
 CC ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: BRAIN, SPINAL CORD, GASTROINTESTINAL TRACT,
 CC SPLEEN AND LEG MUSCLE. IS NOT DETECTED IN THE HEART, LIVER,
 CC STOMACH, LUNG AND KIDNEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73268; CAAS1716.1; -
DR PIR; S33733; S33733.
DR PDB; 1DD; 11-JUL-96.
DR GCRDB; GCR 0745; -
DR PRAM; PF00001; 7tm_1; 1.
DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; 3D-structure.
FT TRANSSEM 1 41
FT DOMAIN 1 41
FT TRANSSEM 42 63
FT DOMAIN 42 63
FT TRANSSEM 64 76
FT DOMAIN 64 76
FT TRANSSEM 77 98
FT DOMAIN 77 98
FT TRANSSEM 99 115
FT DOMAIN 99 115
FT TRANSSEM 136 136
FT DOMAIN 136 136
FT TRANSSEM 156 156
FT DOMAIN 156 156
FT TRANSSEM 178 207
FT DOMAIN 178 207
FT TRANSSEM 227 227
FT DOMAIN 227 227
FT TRANSSEM 255 255
FT DOMAIN 255 255
FT TRANSSEM 275 292
FT DOMAIN 275 292
FT TRANSSEM 318 362
FT DOMAIN 318 362
FT DISULFID 113 191
FT CARBOHYD 11 11
FT CARBOHYD 26 26
FT CARBOHYD 102 102
FT CARBOHYD 186 186
SQ SEQUENCE 362 AA; 41194 MW; 3713AE31 CRC32;

Query Match 32.3%; Score 628; DB 1; Length 362;
Best Local Similarity 39.4%; Pred. No. 1.8e-37;
Matches 137; Conservative 62; Mismatches 133; Indels 16; Gaps 7;
QY 4 TESSLSLGLSGPGSGSEVELDCWFEDKFIILPVSAYAVVLGGLNAPTLMEIFR 63
DB 13 TQPELLAG-GWAGNATKCSLT---KTGFQFYLLPYIYLIVFTGLGNSVAIMPEVH 68
QY 64 LRPMDATATYVHNLALSDTLVLSPLLIYYAAHNRPFCTEICKYRFLPYNNLYCSV 123
DB 69 MRPMGSIYVEMFNALADFLYVLTLPALIFEFNKTDMIFGDVCKLORFETHVNLGSI 128
QY 124 LFLNCISVHRVYLGICHPRLALRMGRPRLAGLLCLAVMLVYVAGVCLPNLEFVTS- 182
DB 129 LFLNCISVHRVYLGICHPRLALRMGRPRLAGLLCLAVMLVYVAGVCLPNLEFVTS- 188
QY 183 VLCHDTTRPEEFHDIYHSSAVMGLFVPCLTVLVYGLMARRL-YQPLPGSAQSSRL 241
DB 189 ITCYDTADELRISYFYNSMCTVFMFCIPRIYVILGCGLYKALIKLDL---NSPLRR 245
QY 242 RSLFTIVLVTVAVCVPHITRTIYLAARL---LEADCVLINVVYKVTPLASAN 298
DB 246 KSLIVLIVLVFAVSLVPHFMKTLNLRALFDQTPQMCFAFNDRKYATQVTRGLASLN 305
QY 299 SCLEPVLVYLLTGDKYRROL---RQLCGGGRPPORTAASALVSLPE 342
DB 306 SCVPLIFLAGDTFERRRLSRATKSSRSRSEPNVQSSSEMTLILTE 353

RESULT 9
ID P2YR_BOVIN STANDARD: PRT: 373 AA.
AC P48042;
DT 01-FEB-1996 (Rel. 33, Created)

```

```

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE P2Y PURINOCEPTOR 1 (AMP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR).
GN P2Y1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-AORTIC ENDOTHELIDM.
RX MEDLINE; 95352058.
RA HENDERSON D.J., ELLIOT D.G., SMITH G.M., WEBB T.E., DAINTY I.A.;
RT Cloning and characterisation of a bovine p2y receptor.
RL Biochem. Biophys. Res. Commun. 212:648-656(1995).
CC -1- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
CC ATP AND ADP.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X87628; CA60958.1; -
DR HSSP; P3496; 1DD.
DR PRAM; PF00001; 7tm_1; 1.
DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT TRANSSEM 1 52
FT DOMAIN 1 52
FT TRANSSEM 53 74
FT DOMAIN 53 74
FT TRANSSEM 75 87
FT DOMAIN 75 87
FT TRANSSEM 88 109
FT DOMAIN 88 109
FT TRANSSEM 110 126
FT DOMAIN 110 126
FT TRANSSEM 127 147
FT DOMAIN 127 147
FT TRANSSEM 148 166
FT DOMAIN 148 166
FT TRANSSEM 167 188
FT DOMAIN 167 188
FT TRANSSEM 189 218
FT DOMAIN 189 218
FT TRANSSEM 219 238
FT DOMAIN 219 238
FT TRANSSEM 239 265
FT DOMAIN 239 265
FT TRANSSEM 266 285
FT DOMAIN 266 285
FT TRANSSEM 286 303
FT DOMAIN 286 303
FT TRANSSEM 304 328
FT DOMAIN 304 328
FT TRANSSEM 329 373
FT DISULFID 124 202
FT CARBOHYD 11 11
FT CARBOHYD 27 27
FT CARBOHYD 113 113
FT CARBOHYD 197 197
SQ SEQUENCE 373 AA; 42287 MW; 5C9B87B5 CRC32;

Query Match 31.9%; Score 621; DB 1; Length 373;
Best Local Similarity 43.4%; Pred. No. 5.7e-37;
Matches 126; Conservative 53; Mismatches 103; Indels 8; Gaps 4;
QY 33 KFFILPVSAYAVVLGGLNAPTLMEIFELRPMDATATYVHNLALSDTLVLSPLTI 92
DB 49 FQYILPAVYIIVFTIFLGLNSVAIMFMFVHMKWSGISYVEMFNALADFLYVLTLPALI 108
QY 93 YYYAANHWPEFGEICKVFLEFYMNLYCSVLFJTCISVHRVYLGICHPRLALRMGRPRLA 152
DB 109 FYEYFNKTDWILFGDAMKQLQRFIFHVNLVYGLSILFLTCSAHYSGVVPPLASLGKRNNA 168
QY 153 GLCLAVLVYVAGVCLVNLFEVITS-NKGTIVLCHDTTRPEEFHDIYHSSAVMGLFV 211
DB 169 VYISVLVWLIVVAGISPLIFYSGIGIRKNKTIITCYDTSDELYRSFYIYSMCTVAMFCV 228
QY 212 PCULTVLCVGLMARRL-YQPLPGSAQSSRLSRJTAVALTVAVCVFPHITRTIYL 270

```

```

Oy      271 ARL---LEADCRVLNVVYKVTBPLASNSCLDPVLLLGDKRRRL 317
         ||| | | | | | | | | | | | | | | | | | | | :||| :|
Db      286 ARLDQTEPMCAFNDRVATYATVGTGSLASNCSDPILFYLAGDFTFRRLL 335

RESULT   10
P2YR_RAT ID P2YR_RAT STANDARD: PRT: 373 AA.
P49651:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE P2Y PURINORECEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR).
GN P2YX1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=INSULINOMA;
RX MEDLINE: 95298025.
RA TOKUYAMA Y., HARA M., JONES E.M.C., FAN Z., BEIL G.I.:
RL "Cloning of rat and mouse p2y purinoreceptors."
RL Blochem. Biophys. Res. Commun. 211:211-218(1995).
CC -1- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
CC ATP AND ADP. IN PANCREATIC ISLETS, MAY MEDIATE SOME OF THE EFFECTS
CC OF EXTRACELLULAR ATP ON INSULIN SECRETION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MUSCLE, HEART, LIVER, KIDNEY,
CC LUNG, BRAIN, SPLEEN, BUT NOT IN TESTIS.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U22830; AAA91303.1; -.
DR HSPR: P34996; IDDD.
DR GCRDB: GCR_144; -.
DR PEM: PF00001; 7tm_1; 1.
DR PROSITE: PS00237; G-PROTEIN RECEPTOR. 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT TRANSSEM 1 52 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 53 74 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 75 87 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 88 109 2 (POTENTIAL).
FT TRANSSEM 110 126 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 127 147 3 (POTENTIAL).
FT TRANSSEM 148 166 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 167 188 4 (POTENTIAL).
FT TRANSSEM 189 218 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 219 238 5 (POTENTIAL).
FT TRANSSEM 239 265 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 266 285 6 (POTENTIAL).
FT TRANSSEM 286 303 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 304 328 7 (POTENTIAL).
FT TRANSSEM 329 373 CYTOPLASMIC (POTENTIAL).
FT DISULPID 124 202 BY SIMILARITY.
FT CARBOHYD 11 11 POTENTIAL.
FT CARBOHYD 27 27 POTENTIAL.
FT CARBOHYD 113 113 POTENTIAL.
FT CARBOHYD 197 197 POTENTIAL.
SQ SEQUENCE 373 AA; 42321 MW; FAF215F8 CRC32;

```

```

Best Local Similarity 43.1%, Pred. No. 1,3e-36; Indels 8; Gaps 4;
Matches 125; Conservative 54; Mismatches 103;

QY      33 EKFIILPVSYAVVEVYLGLNAPLIMLEIFRLRPWDATATYTHFLASDLYLSPLLI   92
       11 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      49 FOEYLLPVAIVYLVEIFIGELGNSVAIMVFEMHKPSGISGYWFNALADLVLTLPALI  108
       11 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      93 YYVAAHNNMPGETELCKRVRELFYNWLCVSYLEFICISVHRYLGLCHPLRALRMGRPRLA  152
       11 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     109 FYENKTDWIMIGDVACQKQRPIEFHNKLGSILFETICISAHRSGVAYPLKSLGRUKRNA  168
       11 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY     153 GLTCLAVWLVAAGCIVPLPLFEVTTS-NKGTVVLCDHTRPEEDFYHESSAVMGLEFGV  211
       11 : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     169 IYVSIVLYGMIIYVAISPILFYESGCIRKNKYTCDSIDSEYLRSEFYISMCTTYAMCCI  228
       11 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY     212 PCLVTLVCYGLMARLL-YQPLPGSNOSSSRLSRTTIYAVLVTFPAVCEVPFHITPTIYYL  270
       11 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     229 PLVILIGCYGLIVRALIYKDLD---NSPLRRKSITVILVTFVASYIPFHVMTMLNR  285
       11 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      21 ARL---LEADCRLNTIVNVVKYTRPLASANSCLDPVLYLLTGDKYRKQL  317
       11 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     286 ARLDQEPMECDPFNDRYATYQVYRGLASLNSCVDPILYFLAGDTFRRL  335
       11 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 11
P2YR_MOUSE STANDARD: PRT; 373 AA.
AC P49650;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE P21 PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR).
GN P2RY1..
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INSULINOMA;
RX MEDLINE; 95298025.
RA TOKUYAMA Y., HARA M., JONES E.M.C., FAN Z., BELL G.I.;
RT "Cloning of rat and mouse p2y purinoceptors."
RL Biochem. Biophys. Res. Commun. 211:211-218(1995).
CC -1- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
CC ATP AND ADP. IN PANCREATIC ISLETS, MAY MEDIATE SOME OF THE EFFECTS
CC OF EXTRACELLULAR ATP ON INSULIN SECRETION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on tests
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U23829; AAA91302.1; -.
DR HSPB: P34996; 1DDO.
DR GCRDB: GCR_1703; -.
DR MGJ: MGI:105049; P2RY1.
DR PFAM: PF00001; 7tm_1; 1.
DR PROSITE: PS00237; G-PROTEIN RECEPTOR: 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 52
FT TRANSSEM 53 74
FT TRANSSEM 75 87
FT TRANSSEM 88 109
FT DOMAIN 110 126
FT TRANSSEM 127 147
FT DOMAIN 148 166
FT TRANSSEM 167 188
FT DOMAIN 189 218
FT EXTRACELLULAR (POTENTIAL).

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC DR EMBL; D63665; BAA09816.1; -
 CC DR HSSP; P34996; IDDD.
 CC DR GCRB; GCR1520; -
 CC DR PFAM; PF00001; 7tm.1; 1.
 CC DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; FALSE_NEG.
 CC KM G-protein coupled receptor; Transmembrane; Glycoprotein.
 CC FT DOMAIN 1 27
 CC FT TRANSMEM 28 48
 CC FT DOMAIN 49 62
 CC FT TRANSMEM 63 83
 CC FT DOMAIN 84 101
 CC FT TRANSMEM 102 122
 CC FT DOMAIN 123 144
 CC FT TRANSMEM 145 165
 CC FT DOMAIN 166 194
 CC FT TRANSMEM 195 215
 CC FT DOMAIN 216 236
 CC FT TRANSMEM 237 257
 CC FT DOMAIN 258 280
 CC FT TRANSMEM 281 303
 CC FT DOMAIN 304 328
 CC FT DISULFID 99 177
 CC FT CARBOHYD 5
 CC SQ SEQUENCE 328 AA; 36677 MW; 2FA27E5C CRC32;

Query Match 30.2%; Score 586; DB 1; Length 328;
 Best Local Similarity 42.2%; Pred. No. 1e-34;
 Matches 136; Conservative 37; Mismatches 133; Indels 16; Gaps 6;

QY 5 ESSLRLSLGSPGSGSEVELDEWDEDFKLLPVSAVAVFVGLGNATLWLFIFRL 64
 DB 4 DNGITQAPGLP-----TTCVYREDFKLLPVSIVLVGLPLNCVIAQICASR 55
 QY 65 RPDWATATYMFHIALDLTYLVLSPTLIYYAAHNMHFEIGKFEVFLFYNNLYCSVL 124
 DB 56 RFLRSAYVTLNLALADLVACSLPLLIYNARGDHMFGLACLVKFLFYALDHSIL 115
 QY 125 FLVCISVHRIYIGICHPRLALNW---GPRRLAGLICLAVYLVAGCLVFNLEFVTSNKG 181
 DB 116 FLVCISVFORYLIGICHPRLAP--WHRKRGRRAAWVCGVWLVVTAQCLPVAFAATGIQRN 173
 QY 182 TVLCHDTTREPEDFHYHSSAVVGLFGVPCVLTIVCYGLMARL-YQPLP-GSAOSS 239
 DB 174 RYCYDLSPLFLSTRIPYGMALTVIGLFTALLACCYCMARLRQDOPAPVQOER 233
 QY 240 RLRLSLRTIAVLLVFAVAFVCFVHITRTIYVLALE-ADCEVLIIVNVVYKVRPLASAN 298
 DB 234 RSKARAAVVAAYVAFVSLPFLHITKTAIYAVRSTPGVSCVLETFPAAYKGTIPFASAN 293
 QY 299 SCIDPVLYLLTGKRYRQLROL 320
 DB 294 SVLDPILEFTQOKFRQPHDL 315

RESULT 14
 P2Y6_HUMAN STANDARD; PRT; 328 AA.
 AC Q15077; Q15754;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE P2Y PURINORECEPTOR 6 (P2Y6).
 GN P2RY6.
 OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC [1]
 CC RA SEQUENCE FROM N.A.
 CC RA KOMMONT D., PARENTIER M., BOEYNAENS J.M.;
 CC RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 97432828.
 CC RA SOMERS G.R., HAMMET F., WOOLLIATT E., RICHARDS R.I., SOUTHEY M.C.,
 CC RA VENTER D.J.;
 CC RT "Chromosomal localization of the human P2Y6 purinoceptor gene and
 CC RT phylogenetic analysis of the P2Y purinoceptor family.";
 CC RL Genomics 44:127-130(1997).
 CC RN [3]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE-PLACENTA;
 CC RX MEDLINE: 97415792.
 CC RA MAIER R., GLATZ A., MOSBACHER J., BILBE G.;
 CC RT "Cloning of P2Y6 cDNAs and identification of a pseudogene: comparison
 CC RT of P2Y receptor subtype expression in bone and brain tissues.";
 CC RL Biochem. Biophys. Res. Commun. 237:297-302(1997).
 CC RN [4]
 CC RP ERRATUM.
 CC RX MEDLINE: 98069816.
 CC RA MAIER R., GLATZ A., MOSBACHER J., BILBE G.;
 CC RL Biochem. Biophys. Res. Commun. 240:298-302(1997).
 CC CC -1- FUNCTION: RECEPTOR FOR EXTRACELLULAR UDP > UTP > ATP. THE ACTIVITY
 CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE A
 CC PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
 CC CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC DR EMBL; X97058; CAA65770.1; -
 CC DR EMBL; U52464; AAB03572.1; -
 CC DR EMBL; AF007892; AAB80713.1; -
 CC DR EMBL; AF007891; AAB80712.1; -
 CC DR HSSP; P34996; IDDD.
 CC DR MIM; 602451; -
 CC DR PFAM; PF00001; 7tm.1; 1.
 CC DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; FALSE_NEG.
 CC KM G-protein coupled receptor; Transmembrane; Glycoprotein.
 CC FT DOMAIN 1 27
 CC FT TRANSMEM 28 48
 CC FT DOMAIN 49 62
 CC FT TRANSMEM 63 83
 CC FT DOMAIN 84 101
 CC FT TRANSMEM 102 122
 CC FT DOMAIN 123 144
 CC FT TRANSMEM 145 165
 CC FT DOMAIN 166 194
 CC FT TRANSMEM 195 215
 CC FT DOMAIN 216 236
 CC FT TRANSMEM 237 257
 CC FT DOMAIN 258 280
 CC FT TRANSMEM 281 303
 CC FT DOMAIN 304 328
 CC FT DISULFID 99 177
 CC FT CARBOHYD 5
 CC FT CONFLICT 2
 CC SQ SEQUENCE 328 AA; 36429 MW; 29C125CE CRC32;

Query Match 30.1%; Score 586; DB 1; Length 328;
 Best Local Similarity 42.5%; Pred. No. 1.4e-34;

Matches 135; Conservative 37; Mismatches 128; Indels 18; Gaps 6;

QY 10 RSIGLSPGSSSEVLDCEWDEDFKILLPVSAYAVVIGLGNAPLWLFIRLRPMDA 69
 DB 9 QALGLPP-----TTCVRENKOLLPLPPVSAVLAAGLPINICVITQICSRALTR 60
 QY 70 TATYMHIALSDTLVLSPTLLYYAAAHNHPGCEICKFVRFYWNLYGCVLEFIC 129
 DB 61 TAYTINLADLADLLVACSPLLYNNAOGDHMPGDFACLVRELFYANHGSIILEFIC 120
 QY 130 SVARYIGICHPALRW---GRPLAGLGLAVMLVAGSLVNLPEVTSNNGITVLC 186
 DB 121 SFRVYIGICHPALRW---WHRKGRRAAMLVCAVAVLATTQCLPTALPAAGIQNRFTVCY 178
 QY 187 DTRPEEFDHVHSSAVMGLFVPCVLTVCYGLMARLYO---PLPGSAOSSRLRS 243
 DB 179 DLSPPALATFHYMYGALVYIGFLPFAALLACVCLACQDQDAEVAQ-ERGRKA 237
 QY 244 LRTIAYVLYFVAFCEVPEHTITTYTLARLE-ADCRVNIIVVYKTRPLASANSCLD 302
 DB 238 ARNAVVAFAAFSTFLPHITKTAYLAVRSTPGVPCVLEAFAPAAVKGTRPFASANSVLD 297
 QY 303 PVLYLTGDKYRROL 320
 DB 298 PILFTYOKKERRRPHL 315

RESULT 15
 GPRH HUMAN
 ID GPRH HUMAN STANDARD; PRT; 339 AA.
 AC 013304;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PUTATIVE G PROTEIN-COUPLED RECEPTOR GPR17 (R12).
 GN GPR17.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96145150.
 RA RAPPORT C.J., SCHWEICKART V.L., CHANTRY D., EDDY R.L. JR., SHOWS T.B.,
 RA GODISKA R., GRAY P.W.;
 RT "New members of the chemokine receptor gene family.";
 RL J. Leukoc. Biol. 59:18-23(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HIPPOCAMPUS;
 RX MEDLINE; 98181695.
 RA BLAESIUS R.H., WEBER R.G., LICHTER P., OGILVIE A.;
 RT "A novel orphan G protein-coupled receptor primarily expressed in the
 brain is localized on human chromosome band 2q21.";
 RL J. Neurochem. 70:1357-1365(1998).
 RT
 CC -1- FUNCTION: ORPHAN RECEPTOR. POTENTIAL CHEMOKINE RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U33447; AAB16746.1;
 DR EMBL: Z94154; CAB08107.1;
 DR HSSP: P34996; 1DD.
 DR GCRDB; GCR_1968;
 DR GCRDB; GCR_2597;
 DR MIM: 603071;

DR PFAM: PF00001; 7tm_1. 1.
 DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 36
 FT TRANSMEM 37 57
 FT DOMAIN 58 64
 FT TRANSMEM 65 85
 FT DOMAIN 86 105
 FT TRANSMEM 106 126
 FT DOMAIN 127 147
 FT TRANSMEM 148 168
 FT DOMAIN 169 195
 FT TRANSMEM 196 216
 FT DOMAIN 217 232
 FT TRANSMEM 233 253
 FT DOMAIN 254 280
 FT TRANSMEM 281 301
 FT DOMAIN 302 339
 FT DISULFID 104 181
 FT CARBOHYD 14 14
 FT CARBOHYD 176 176
 FT CARBOHYD 254 254
 SQ SEQUENCE 339 AA; 37826 MW; CFA6CB30 CRC32;

Query Match 26.2%; Score 510; DB 1; Length 339;
 Best Local Similarity 37.1%; Pred. No. 3, 1e-29;
 Matches 119; Conservative 52; Mismatches 136; Indels 14; Gaps 7;

QY 27 CWDEDEKFKILLPVSAYAVVIGLGNAPLWLFIRLRPMDATATYMHIALSDTLVYL 86
 DB 23 CGGETPLENMLFASFYFLDLPLALVENTILALWFINDHSGTFANFIMHLAVADLSVYL 82
 QY 87 SPTLLYYAAAHNHPGCEICKFVRFYWNLYGCVLEFICISVARYIGICHPALRW 146
 DB 83 VLPTRLVYHSGNHPGCEICKFVRFYWNLYGCVLEFICISADRFALIVAPKSKL 142
 QY 147 GRPLAGLGLAVMLVAGSLVNLPEVTSNNGITVLCDDTRPEEFDHVHSSAVMG 206
 DB 143 RPLVYHHLACAFILVAVVAVMAAPLVSPOITVQTNHVVCLQLR-ERASHHALVSLAV-- 200
 QY 207 LTFGVPCVLTVCYGLMARLYOPLPGSAOSSRLRSRTIAYVLYFVAFCEVPHITRT 266
 DB 200 -ATFPFITTVCYLLIITSRLGL--RVKRLKTAVRALIVLALFVCEVPIYHNS 256
 QY 267 IYVL-ARLEADCRVNIIVVYKTRPLASANSCLDPVLYLTGDKYRROL-ROLG-- 323
 DB 257 VYVLYHRSHGASCATQRIILANRITSCLTSLGALDPIMYFVAKFRHALCNLLGKR 316
 QY 323 --GGRP--QPRPAASSLAVS 339
 DB 317 LKGPSPSFEKGKNESSLSAKS 337

Search completed: December 29, 1999, 20:19:11
 Job time: 2130 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 1999, 20:41:04 ; Search time 16.07 Seconds

(without alignments)
1391.297 Million cell updates/sec

Title: us-09-077-173a-2

Perfect score: 1946

Sequence: 1 MASTESSLLSLGSLSPGPGS.....CRMAATPDSSCSTPRADRF 365

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database : SPTREMBL_11.*

Word size : 0

Number of hits that pass the threshold : 199794

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1593	81.9	361	11	035811
2	1137.5	57.9	374	13	057466
3	599	30.8	328	13	093361
4	450.5	23.2	370	4	015132
5	400	20.6	353	13	P79960
6	374	19.2	415	4	015185
7	368	18.9	383	13	042324
8	356.5	18.3	359	6	077590
9	354	18.2	339	4	000398
10	352	18.1	303	13	P70058
11	352	18.1	373	13	057585
12	351	18.0	383	12	089609
13	349	17.9	333	6	018951
14	348.5	17.9	381	4	095853
15	342	17.6	371	4	043190
16	337.5	17.3	380	6	097666
17	334	17.2	342	13	093239
18	332	17.1	356	6	097665
19	332	17.1	353	13	093247
20	331.5	17.0	396	11	088634
21	329.5	16.9	338	13	09YGC3
22	328	16.9	352	6	077488
23	328	16.9	357	13	042445
24	327.5	16.8	360	4	060835
25	325	16.7	385	4	076067

ALIGNMENTS

```
RESULT 1
ID 035811 PRELIMINARY; PRT; 361 AA.
AC 035811;
DR 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE G-PROTEIN COUPLED RECEPTOR.
GN P2Y4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LIVER;
RA BOGDANOV Y.D., WILDMAN S., KING B.F., BURNTOCK G.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;
RX MEDLINE; 98421785.
RA WEBB T.E., HENDERSON D., ROBERTS J.A., BARNARD E.A.;
RT "Molecular cloning and characterization of the rat P2Y4 receptor.";
DR EMBL; Y14705; CAA75007.1; -
DR EMBL; Y11433; CAA72241.1; -
DR PIRAM; PF00001; 7tm_1; 1.
KW G-protein coupled receptor.
SQ SEQUENCE 361 AA; 40893 MW; D1882F0C CRC32;

Query Match 81.9%; Score 1593; DB 11; Length 361;
Best Local Similarity 82.7%; Pred. No. 2e-131;
Matches 301; Conservative 20; Mismatches 39; Indels 4; Gaps 1;

OY 1 MASTESSLLSLGSLSPGPGSEVELDCWFDEKFIILPVSYAVVVLGLNAPTLLWF 60
DB 1 MTSASLLFTSLGSPSSGDC---DCRFNEEFKFIILPMASVAVFVGLNAPTLLWF 56
OY 61 IFRAPMATAATYMHIALSDTLVYLSPTLIYYAAHHNHPFGIEIKFVRFLEFYNLY 120
DB IFRAPMATAATYMHIALSDTLVYLSPTLIYYAAHHNHPFGIEIKFVRFLEFYNLY 116
OY 57 LFRAPMATAATYMHIALSDTLVYLSPTLIYYAAHHNHPFGIEIKFVRFLEFYNLY 116
DB LFRAPMATAATYMHIALSDTLVYLSPTLIYYAAHHNHPFGIEIKFVRFLEFYNLY 116
OY 121 CSVLEFICISVHRVYIGICHPRLALRMGRPRLAGLCLAVLTVAGCLVPNFEVYTSNGK 180
DB 117 CSVLEFICISVHRVYIGICHPRLALRMGRPRLAGLCLAVLTVAGCLVPNFEVYTSNGK 176
OY 181 TTVLCHDITRPEEDHDYVHFSSAVMGLFGVPCLTIVLCYGLMARLYQPLPGSAOSSSR 240
```

```

DB 177 TLTLCHDTLPEEFDAHYVFFSSAVMYLLGLPLTLVCGLMARLRLPLPAGSSSR 236
OY 241 LNSLRTIAVLVFAVCFVPHITRTIYLLARLEADCRVLINVVYKTRPLASANGC 300
DB 237 LNSLRTIAVLVFAVCFVPHITRTIYQARLQADCHVNLVNVYKTRPLASANGC 296
OY 301 LBPVLYLLTGDKYRRQLRQLCGGKPOPTAASLAVSLPEDSSCRMAATPODSSCSTP 360
DB 297 LBPVLYLTGDKYRRQLRQLCGGKPOPTAASLAVSLPEDSSCRMAATPODSSCSTP 356
OY 361 RADR 364
DB 357 EGDR 360

RESULT 2
O57466 PRELIMINARY: PRT: 374 AA.
AC 057466:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE G PROTEIN COUPLED P2Y NUCLEOTIDE RECEPTOR.
GN TP2Y.
OS Meleagris gallopavo (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Meleagrididae; Meleagris.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD:
RX MEDLINE: 98086419.
RA BOYER J.L., WALDO G.L., HARDEN T.K.;
RT "Molecular cloning and expression of an avian G protein-coupled p2y
RT receptor";
RL MOL. Pharmacol. 52:928-934(1997).
DR EMBL: AF031897; AAC03039.1; -.
DR PFM: PF00001; 7tm_1; 1.
SQ SEQUENCE 374 AA; 42594 MW; 5320428C CRC32;

```

```

Query Match 57.9%; Score 1127.5; DB 13; Length 374;
Best Local Similarity 59.3%; Pred. No. 7.8e-91;
Matches 208; Conservative 56; Mismatches 70; Indels 17; Gaps 4;

OY 9 LRSGLSP-----GPSSEVELDCWDEDFKILLPVSYAVFVLGGLNAPTLM 58
DB 5 VAMESLAPPTPTPTWLGNTTAAAEAKCVNEEFKILLPISYGVFVGLPLNSMAM 64
OY 59 LFFRLRPMDATATYMHFALSDTLVLSPLTIYYAAHNMHPGTEICKVRFLEFYN 118
DB 65 IFVSMRRMNNATTITMFNLAISDTLVPSLPLVYYADRNMWPGKVCYKTRLEFAN 124
OY 119 LYCSVLFITCISVHRYLGICHPRLARMGRPRLAGLCLAVLVVAGCLVPLNLFETSN 178
DB 125 LYSSTLFLTCISVHRYMGICHPIRSLKWKYTKHARLICGVWLVTICILPMLIEVTS 184
OY 179 KGTVLCGDTTRPEEDHVVHSSAVMGLLEFVPCVLTVYCGLMARLIQ---PLPSGA 235
DB 185 KDNSTLCHDTKPEEFDAHYVFFSSSIMLLFGLPLVYVCCLMARLCKRSPSPSR 244
OY 236 QSSRLRLRTIAVLVFAVCFVPHITRTIYLLARLEADCRVLINVVYKTRPLA 295
DB 245 VPSYKRRIMKIIVLYVFAICFVPHITRTIYYSRFOACQDILNFTYKTRPLA 304
OY 236 SANSCLDPVLLTGDKYRRQLRQLCGGKPOPTAAS-LALVSLPEDSS 345
DB 305 SINSLDPILYFMAGDKYRGLRR---GAAGRPRVPVPTSLALVSPVSDS 352

```

```

RESULT 3
O93361 PRELIMINARY: PRT: 328 AA.
ID O93361

```

```

AC O93361:
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE G PROTEIN-COUPLED RECEPTOR P2Y3.
GN P2Y3.
OS Meleagris gallopavo (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Meleagrididae; Meleagris.
RN [1]
RP SEQUENCE FROM N.A.
RA LI Q., OLESKY M., PALMER R.K., HARDEN T.K., NICHOLAS R.A.;
RT "Evidence that the p2y3 receptor is the avian homologue of the
RT mammalian p2y6 receptor.";
RL MOL. Pharmacol. 0:0-0(1998).
DR EMBL: AF069555; AAC23863.1; -.
DR PFM: PF00001; 7tm_1; 1.
SQ SEQUENCE 328 AA; 37594 MW; 1E0F2067 CRC32;

```

```

Query Match 30.8%; Score 599; DB 13; Length 328;
Best Local Similarity 42.3%; Pred. No. 7.9e-45;
Matches 127; Conservative 44; Mismatches 121; Indels 8; Gaps 4;

OY 27 CWDEDFKILLPVSYAVFVLGGLNAPTLMFLFRLRPMDATATYMHFALSDTLVYL 86
DB 13 CFQDEFFQVLLPVTYSVFLGLPLNNAVIGQWLAKRALTRITTYMLNLTADLLVC 72
OY 87 SLPLTIYYAAHNMHPGTEICKVRFLEFYNLYCSVLFITCISVHRYLGICHPRLARM 146
DB 73 SLPLTIYNTQDKWMPFGDETKFVRFQFYTLHGSILFLTCISVORWIGICPLAS--W 130
OY 147 GR---PRLAGLCLAVLVVAGCLVPLNLFETTSKGTIVLCHDTTRPEEDHVFSSA 203
DB 131 HKRKKRLTWLVCAVWEIYVIAQCLPTEVFASTGTQRNRTVYDLSPPDRSASYPGIT 190
OY 204 VMLGLEVPCLTVYCYGMARLYQ---PLPSAQSSTLSRLRTIAVLVFAVCFVPE 261
DB 191 LTTTFLLPFAILTCYSMARILQKDELGLAVHKKKDKRVRMIIIVYVTSISFPE 250
OY 262 HITRTIYLLARLEA-DCRVLINVVYKTRPLASANSCLDPVLYLLTGDKYRRQLROL 320
DB 251 HHTKTIYIVSSPPLPCTQLQAFALAKCTRPFSMNSVLDPLIFFTQKRFRSTRYL 310

```

```

RESULT 4
O15132 PRELIMINARY: PRT: 370 AA.
ID O15132
AC O15132:
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE P2Y5-LIKE RECEPTOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 97366605.
RA JANSSENS R., BOEYNAEMS J.M., GODART M., COMMUNI D.;
RT "Cloning of a human heptahelical receptor closely related to the P2Y5
RT receptor.";
RL Biochem. Biophys. Res. Commun. 236:106-112(1997).
DR EMBL: AF005419; AAB66322.1; -.
DR PFM: PF00001; 7tm_1; 1.
SQ SEQUENCE 370 AA; 41861 MW; A42AB01D CRC32;

```

```

Query Match 23.2%; Score 450.5; DB 4; Length 370;
Best Local Similarity 33.4%; Pred. No. 7.9e-32;
Matches 102; Conservative 67; Mismatches 131; Indels 5; Gaps 5;

OY 14 LSPGGSSEVELDCWDEDFKILLPVSYAVFVLGGLNAPTLMFLFRLRPMDATATY 73

```


Db 18 LRPRLNATRNATCTIVDDSKYKLNKGAIVSVFLGLITNVSLSFVCFMRKMRSEAIIF 77
 QY 74 MEHLASDLYLVSLPTLLIYYAAHNHMPGTEICKFVRLEFVNLVLCVSLFLTCISVHR 133
 Db 78 ITMLAVSDLLFVCTLPFKI-FYFNHMPGDLICKISGAFLTNITIGSLMFLTCISVDR 136
 QY 134 YLGICHPRLALNMGPRRLAGLCLAVN-LVAGCLVPLEFVTTSNKGTVLCHDTTPE 192
 Db 137 FLAIYVPFRSRTIRTRNSAIVCAGWILVLSGSIASLSTNNATT-CEEGSKR 195
 QY 193 EEPHYHVFSSAVVGL-FGVPCLVTVLCYGLAMRLYQPLPGSAOSSRLSLRTAVVL 251
 Db 196 VMTYLSKTIIFLEVGFILPLINWSSSVYRLTRKPAITLSQIGTKKKVLMITVHM 255
 QY 252 TVFAVCFEPHITRTIYIYARLE-ADCRVLINVVYKTRPLASNSCLDPVLYLTG 310
 Db 256 AVFVVCVFVYNSVLFILALVRSQAITNCFLEFRAKIMYPITLCLATLNCCEFDPIYFTL 315
 QY 311 DKYRR 315
 Db 316 ESFOK 320

RESULT 5
 ID P79960 PRELIMINARY; PRT: 353 AA.
 AC P79960;
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE MESNCNHYME-ASSOCIATED SERPENTINE RECEPTOR.
 GN X-MSR.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97109511.
 RA DEVIC E., PAQUEREAU L., VERNIER P., KNIBIEHLER B., AUDIGIER Y.,
 RT "Expression of a new G protein-coupled receptor X-msr is associated
 RL with an endothelial lineage in Xenopus laevis."
 CC Mech. Dev. 59:129-140(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 DR EMBL: X93045; CA63612.1; -.
 DR PFIAM: PF00001; 7tm_1.1.
 DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein.
 SQ SEQUENCE 353 AA; 40303 MW; 3354C3AE CRC32;

Query Match 20.6%; Score 400; DB 13; Length 353;
 Best Local Similarity 33.7%; Pred. No. 1.9e-27;
 Matches 106; Conservative 59; Mismatches 130; Indels 20; Gaps 10;

QY 30 DDDDFILLPVSAVAVYGLGNAITLMLFELRLRPMATA-TYFHLASDLYLVLS 88
 Db 30 DDDDFILLPVSAVAVYGLGNAITLMLFELRLRPMATA-TYFHLASDLYLVLS 88
 Db 30 DDDDFILLPVSAVAVYGLGNAITLMLFELRLRPMATA-TYFHLASDLYLVLS 88
 QY 89 PTLIYYAAHNHMPGTEICKFVRLEFVNLVLCVSLFLTCISVHRIGLCHPLRALMGR 148
 Db 89 PTLIYYAAHNHMPGTEICKFVRLEFVNLVLCVSLFLTCISVHRIGLCHPLRALMGR 148
 Db 89 PTLIYYAAHNHMPGTEICKFVRLEFVNLVLCVSLFLTCISVHRIGLCHPLRALMGR 148
 QY 149 PRLAGLCLAVN-LVAGCLVPLEFVTTSNKGTVLCHDTTPEERDHH---VHR 200
 Db 149 PRLAGLCLAVN-LVAGCLVPLEFVTTSNKGTVLCHDTTPEERDHH---VHR 200
 Db 148 SRPSITVSLAVITLWGLSLALPSLIIRDTRVBNNTICDIDFSGVSKENENWIGLST 207
 QY 201 SSAVGLGVPCLVTVLCYGLAMRLYQPLPGSAOSSRLSLRTAVVLTVFAVC 257
 Db 208 LTVBGFLL-LPRLMTIIFYCFIGKVTMHPVFLKKEOKKRL-LKIIITLVVFAIC 263

QY 258 FVPEHITRTIYIYARLE-LEADCRVLINVVYKTRPLASNSCLDPVLYLTGDKYRR 315
 Db 264 WLPFHILKTHIFLDMGLFELSCSTQNIIVSLRPHATCLAVNSCLNPFILAFDLFRS 323
 QY 316 QLRLOGGGKPOPT 330
 Db 324 QCFEFPFVKSPPT 338

RESULT 6
 ID 015185 PRELIMINARY; PRT: 415 AA.
 AC 015185;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE G PROTEIN-COUPLED RECEPTOR CKR-12.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA GUTIERREZ J., VARONA R., ZABALLOS A., LIND P., MARQUEZ G.;
 RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL: Z79783; CAB02143.1; -.
 DR PFIAM: PF00001; 7tm_1.1; -.
 SQ SEQUENCE 415 AA; 45608 MW; EB0DF045 CRC32;

Query Match 19.2%; Score 374; DB 4; Length 415;
 Best Local Similarity 30.6%; Pred. No. 4.1e-25;
 Matches 114; Conservative 57; Mismatches 155; Indels 46; Gaps 11;

QY 6 SSILRSLSGSPGSSSEVELDCM-----FDEDFKILLPVSAVAVYGLGNAITL 58
 Db 64 ALLENFSSSYDGENESDSCSPCPQDFSLNDRALPALYSILFLGLGNGAVRA 123
 QY 59 LFIERLPMDATYTFHIALSDLYLVSLPTLLIYYAAHNHMPGTEICKFVRLEFYNN 118
 Db 124 YLSRRTALSDTDFLHLAVADTLVTLTP-LMAVDAVQVWFSGCKVAGALFNIN 181
 QY 119 LYCSVFLFCISVHRIGLCHPLRALMGRPRLAGLCLAVNLYVAGCLVPLNFFTTSN 178
 Db 182 FYAGALLILCISFDRNLIVHATOLYRPPARVYTLCLAWGLCLFLALPDIFISAH- 241
 QY 179 KGTTVLCHD-----TRPEEPDHYVHSSAVVGLFG--VPCLVTVYCG-LMARLYOP 230
 Db 241 -----HDERLNATGQYNFQVGRATLRVQLVAGFLPLLVAYCTAHILAVLV-- 292
 QY 231 LFGSAOSSRLSLRTIAVLTVFAVCFVPHITRTIYIYARLE-LEADCRVLINVVY 288
 Db 292 -----SRGQRRRLAMRLVVVVVAVAFALCMPTPHLVVLDIMDLGALARNGRESRDVAK 347
 QY 289 KYTRPLASNSCLDPVLYLTGDKYRRQ-----LROLGGGKPOPTTAASSLALSLPED 343
 Db 348 SVTSGGLYWHCCINPLLVAFVGVKFRERMMDLRLGCPNORGLOROPSSS-----RRD 401
 QY 344 SSCRMATPODS 355
 Db 402 SS--WSETSEAS 411
 RESULT 7
 ID 042324 PRELIMINARY; PRT: 383 AA.
 AC 042324;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE MU-OP10ID RECEPTOR.
 OS Catostomus commersoni (White sucker).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;


```

Db      92 YIYSH-HMPFORALCLCEFLKYLNMYASICFLTCSIGRCPEFLKPFARBMKRRYDV 150
OY      153 GLGLAVLVYVAGCLVPNLFVYTT---SNKGTIVLCHDTRPEEFDHVHFSSAVMGL- 209
Db      151 G-ISAIMIVYVGTACLPFPLRSTDLNNKNS---CFADGYKQMN-----AVALYGMIT 200
OY      209 -----FGVPCVLTVLCYGLMARLQPLPGSAGSSRLRSKTIANVLVFAVCEVPEFH 262
Db      201 VAEIAGEVLPVITIIAMCTMKTTISLKHOPPMAGFQISEROKALRYVMCAAFVFCFEPYH 260
OY      263 IRTTIVYLAR-LLEADCRVLNIYVNYKYTRPLASANSCLDPVLYLLTGSKYRQL 317
Db      261 INFITYMYKHEIISSCPYRYRIALYFHPCLCLASLCCLDLPILYTFMASEFRDQL 316

RESULT 10
P70058
AC      P70058. PRELIMINARY: PRT: 303 AA.
DT      01-FEB-1997 (TRENBLrel. 02, Created)
DT      01-FEB-1997 (TRENBLrel. 02, last sequence update)
DT      01-MAY-1999 (TRENBLrel. 10, last annotation update)
DE      ANGIOGENSIN RECEPTOR RELATED PROTEIN.
GN      XANGIOI.
OS      Xenopus laevis (African clawed frog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC      Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae; Xenopodinae;
OC      Xenopus.
RN      [1]
RP      SEQUENCE FROM N.A.
RA      SARA M.S., OAKES J.A., MILES R.R.;
RL      Submitted (SEP-1996) to the EMBL/Genbank/DDI databases.
CC      1- SUBCELLULAR LOCATION: INTERAL MEMBRANE PROTEIN (BI SIMILARITY).
CC      1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR      EMBL; U72029; AAB17004.1; -.
FW      PFM; PF00001; 7tm_1; 1.
KW      PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
KW      G-protein coupled receptor; Transmembrane; Glycoprotein.
SEQUENCE 303 AA: 34068 MW; F04BBF0B CRC32;

```

DR EMBL; U20824; AAC13866.1; -
 DR EMBL; U20824; AAC13788.1; -
 DR PFM; PF00001; 7tm_1.1;
 SO SEQUENCE 383 AA; 43667 MM; 60F5BPD4 CRC32;

Query Match 18.0%; Score 351; DB 12; Length 383;

Best Local Similarity 26.3%; Pred. No. 3.8e-23;
 Matches 97; Conservative 74; Mismatches 150; Indels 48; Gaps 9;

QY 1 MASTESSLSRSL-----GLSPGSGSEVELDCWFEDEFK-----FILPYS 41
 DB 19 MTNFTSLTSVATTIASLVPSTNSSEDDYDDLDVDESAPCYKSDPTRLAAQVAPAL 78
 QY 42 YAVVVLGLGNAPLMLFIFRLRPMDATATMFLASDITLYSLPLIYYIAAHNM 101
 DB 79 YLVLVFLGLGLVILVIVIRYMKIKNLMLMLAISDLFLTLPLFMHYIYIMYHM 138
 QY 102 PFGTICKFVRLFYVNYLCSVLFCTSVHRYLGICHPRLRMGRPRLAGLCLAVWL 161
 DB 139 TFGISLCLKLGKVCYMSLXSOVFCITILTVDRYLAVTALRFRYTCGIVTCVTF 198
 QY 162 VVAGCLVPLNEFVTTSNKGTIVLCHDTRPEEDHY--VHFSAVMGLLEGVCLVTLV 218
 DB 199 LAGLLSLPEFFHGHQDNGRYQC-DPYYPEKSTVWRAHAKYIM-LSLLPLIMAY 256
 QY 219 CYGLMARLYQPLPSAASSSRSLRLTIAVLYVAVCFVPHITRTI-YYLARLLEAD 277
 DB 257 CYVYIRRLR-----RPSKKRYKAIKRLFVIMAVYFVFWPTNYIVLSTFHAHLNQ 311
 QY 278 CRVLIVVVVYVTRPRLASNSCLDPVLYLNGDKYRQLRG-----LC----- 322
 DB 312 CALSNLDMALLITVTAVATHCCINPVIAFVEKFRHLHFHTYVAIYICKYIPLS 371
 QY 322 --GGGKPP 328
 DB 372 GDGEKEGP 380

RESULT 13
 ID 018951 PRELIMINARY; PRT; 133 AA.
 AC 018951;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE P2U/P2Y2 RECEPTOR (FRAGMENT).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA CHANG A.S., CHANG S.M., SCHILLING W.P.;
 RL Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF005153; AAB62543.1; -
 DR PFM; PF00001; 7tm_1.1;
 FT NON_TER 1
 FT NON_TER 133
 SO SEQUENCE 133 AA; 15006 MM; 190A8B1F CRC32;

Query Match 17.9%; Score 349; DB 6; Length 133;
 Best Local Similarity 51.1%; Pred. No. 1.9e-23;
 Matches 68; Conservative 23; Mismatches 40; Indels 2; Gaps 1;

QY 129 ISVHRYLIGICHLRLRMGRPRLAGLCLAVLVVAGCLVPLNFVTTSNKGTIVLCHDT 188
 DB 1 ISVHRYLIGICHLRLRMGRPRLAGLCLAVLVVAGCLVPLNFVTTSNKGTIVLCHDT 60
 QY 189 TRPEEDHYVHFSAVMGLLEGVCLVTLVCGLMARLYQPLPSAASSSRSL--RSLRT 246
 DB 61 SARELFSHFVAVSSVMLSLTFAAPRAVILVYGLMARLRLKRAYGTSGGLPRAKKRSVRT 120

QY 247 IAVLTVFAVCFV 259
 DB 121 IAVLTVFVLCEL 133

RESULT 14

ID 095853 PRELIMINARY; PRT; 381 AA.

AC 095853;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE ORPHAN G PROTEIN-COUPLED RECEPTOR.
 GN GPR34.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 99156852.
 RA MARCHESE A., SANZDARGO M., NGUYEN T., CHENG R., HENG H.H., NOWAK T.,
 RA IM D.S., LYNCH K.R., GEORGE S.R., O'DOWD B.F.;
 RT "Discovery of three novel orphan G-protein-coupled receptors."
 RL Genomics 56:12-21(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 DR EMBL; AF118670; AAD17248.1; -
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR; 1.
 KW Receptor; G-protein coupled receptor; Transmembrane; Glycoprotein.
 SO SEQUENCE 381 AA; 43860 MM; 07C967ED CRC32;

Query Match 17.9%; Score 348.5; DB 4; Length 381;
 Best Local Similarity 27.1%; Pred. No. 6.3e-23;
 Matches 95; Conservative 68; Mismatches 156; Indels 31; Gaps 6;

QY 27 CWDEDEKFIPLVSYAVVFLVGLNAPLMLFIFRLRPMDATATYMFHLASDITLYVL 86
 DB 46 CPMEKELSTVLTTSYSVIFVGLVGNILAYFLGIRKRNISQIYLLNVAIDLILFI 105
 QY 87 SLPLIYYIAAHNMHPRETICKFVRLFYVNYLCSVLFCTISVHRYLGICHPRLRLW 146
 DB 106 CLPFRIMHINQNMWTLGVILCKYVGLTFYNNMYISILGFISLDRIKINRSIOOKRA 165
 QY 147 GRPLAGLCLAVLVVAGCLVPLNFVTTSNKGTIVLCHDTRPEEDHYVHFSAA--- 204
 DB 166 ITTKOSIVCCIVMMLALGGLTMIILTLKKGHNSTYCF-----HYRDKHMANGE 216
 QY 204 ----VMGLFVPCVLTLYVCGLMARLYQ-----PLPSAASSSRSLRLTIAVVL 252
 DB 217 AIFNFILVVMFMLFLLIILSIKIKNLRLSKRSKFPNSGVAT--TARNSFIYLI 273
 QY 253 VFVCFVPHITRTIYYIARLLEDCRLNVVNVYKATRLAANSCLDPVLLITSDK 312
 DB 274 IFTICFVPHAFRIYISSLNVSSCYKWEIYHKTNEIMLVLSFNSCLDPVMTFLMSN 333
 QY 313 YRROLQLC---GGGKPPRTAASSLALVSLPDSCKWATPDSSCST 359
 DB 334 IRKIMCOLLFRFQGEPSRSESTSEFRGYSLHDTSV--AVKIQSSSKST 381

RESULT 15

ID 043190 PRELIMINARY; PRT; 371 AA.

AC 043190;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE PURINERGIC P2Y11 RECEPTOR.
 GN P2Y11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA COMMUNI D.; GOVARTS C.; PARMENTIER M.; BOEYNAEMS J.M.;
RL J. Biol. Chem. 0:0-0(1997).
DR EMBL; AF030335; AAB8674.1; -.
DR PFM; PF00001; 7tm_1; 1.
SQ SEQUENCE 371 AA; 40174 MW; 5686F41C CRC32;

Query Match 17.6%; Score 342; DB 4; Length 371;
Best Local Similarity 32.0%; Pred. No. 2.3e-22;
Matches 97; Conservative 44; Mismatches 132; Indels 30; Gaps 7;

QY 29 FDEDEKFLIPVSAVAVFVLGLNAPTLWLF-IFRLRPMDATATYMEHLASDTLYLS 87
DB 23 FQGEFLPIL---VVEFLVAVASNGLALYRFSSIRKORPMHPAVVFSYOLAVSDLLCALT 78
QY 88 LPTLIYYAAHNMWPEFTICKFVRFLEFYNLYCSVLELTCISVHRYLGICHPRLALRWG 147
DB 79 LPPLAALYPPKHRYGEACRLERFLFTCNLGSVIFITCISLNRYLGIVHPEFANSHL 138
QY 148 RPRLAGLLCLAVMLVAGCLVPNLFEVYTSNKGTTVLCHDTRPE-----EFDH---- 197
DB 139 RPKHMAVSAAGVLAALAMPPLTSFSLKRPQAGAGNCVSAREACIKCIGTADHGIAA 198
QY 197 YVHSSAVMGLLPGVPLVTVLCGLMARRLYQPLPGSAOSSRLRSIRTIAVVLTVFAV 256
DB 199 YRAVSLVLAGLGGPLPLTLAAGALGRAVLRSPGWT-VAEKRYAALVAASGVALYAS 256
QY 257 CEVPEHITRTIYYLAR-----LLEADCRVLINVVYKVRPLASANSCLDPV 304
DB 257 SYVPYHMRVLTNDARRKRWSTRCSFADIAQATALELGPYVGTYVMKGLMPLAFVHPL 316
QY 305 LYL 307
DB 317 LYM 319

Search completed: December 30, 1999, 20:49:41
Job time: 517 sec

GenCore version 4.5.
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 03:43:59 ; Search time 727.51 Seconds
(without alignments)
6733.258 Million cell updates/sec

Title: US-09-077-173A-1
Perfect score: 1429
Sequence: 1 AAGGAGCTTGCTAGGGGC.....GCCACGAGCAGGAGATATC 1429

Scoring table: IDENTITY_NUC

Searched: 4089388 segs, 1713965092 residues

Database: EST:*

Word size: 0

Number of hits that pass the threshold: 818776

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*

50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est17:*
66: em_est18:*
67: em_est19:*
68: em_est30:*
69: gb_gss1:*
70: gb_gss2:*
71: gb_gss3:*
72: gb_gss4:*
73: em_gss1:*
74: em_gss2:*
75: em_gss3:*
76: em_gss4:*
77: gb_gss5:*
78: gb_gss6:*
79: gb_gss7:*
80: gb_gss8:*
81: gb_gss9:*
82: em_gss5:*
83: em_gss6:*
84: em_gss7:*
85: em_gss8:*
86: em_gss9:*
87: em_gss10:*
88: em_gss11:*
89: gb_gss10:*
90: gb_gss11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	130.8	9.2	527	63	A1956440	A1956440 UL77C02.Y
2	117.2	8.2	586	69	FR0019937	AL017824 F.rubripe
3	80.2	5.6	544	45	A1344017	A1344017 tcol1e02.x
4	75.4	5.3	542	63	A1942392	A1942392 w077912.x
5	74.2	5.2	466	21	R15256	R15256 Yf89403.r1
6	74.2	5.2	316	31	AA296456	AA296456 EST11038
7	73.8	5.2	261	32	AA331167	AA331167 EST35211
8	70.8	5.0	599	38	AA804531	AA804531 ns28C04.s
9	70.8	5.0	592	64	AL048512	AL048512 DKF2P586M
10	70	4.9	529	29	AA177828	AA177828 m07f02.r
11	66	4.6	650	36	AA611127	AA611127 v061f08.r
12	66	4.6	476	48	A1607213	A1607213 v061f08.y
13	65.8	4.6	815	30	AA205847	AA205847 z450C02.T
14	65.4	4.6	570	51	A1739295	A1739295 z450C02.T
15	62.6	4.4	391	25	N93987	A1739295 w130B03.x
16	62.6	4.4	559	30	AA239704	AA239704 z450C02.T
17	60.4	4.2	249	23	H29103	AA239704 w15C07.r
18	59.2	4.1	274	21	F03704	F03704 HSC27C022.n
19	58.2	4.1	520	49	A1663305	A1663305 UK27C10.Y
20	57.2	4.0	361	21	F07588	F07588 HSC2B071.n
21	57.2	4.0	480	23	H20663	H20663 ym47h08.r1
22	56.4	3.9	733	63	A1959485	A1959485 fd10g03.Y

23	56	3	9	504	37	AA690766
24	55	3	8	456	32	AA354393
25	54.2	3	8	453	25	AA353933 EST62864
26	54.2	3	8	478	37	H97311 EST42110 WH
27	53	3	7	575	63	AA672048 vi07E01.r
28	52.2	3	7	446	44	AI979910 pat. pK00002
29	51.2	3	6	446	44	AI133863 uc36H08.r
30	50	3	5	486	31	AI263552 q120C06.x
31	49.6	3	5	461	50	AA290864 zt19A09.s
32	49.4	3	5	385	23	AI670734 wc28E08.x
33	49.4	3	5	484	46	R91585 yq10A01.r1
34	49.2	3	4	970	69	AI417605 ts80B06.x
35	49.2	3	4	599	69	AL096787 Drosophila1
36	48.6	3	4	997	69	AL012863 F. rubripes
37	48.2	3	4	656	64	AL060767 Drosophila1
38	48.2	3	4	509	69	AL119185 DKFZp761N1
39	47	3	3	595	40	AL012836 F. rubripes
40	47	3	3	1100	70	AA940565 v446E06.r
41	46.6	3	3	243	31	AI106885 Drosophila1
42	46	3	2	371	22	AA298823 EST114454
43	46	3	2	436	38	R80054 y191f07.r1
44	45.4	3	2	781	47	AA773241 at578G04.r
45	45.2	3	2	1101	69	AV051007 AV051007
						AL064052 Drosophila1
						CNS006BP

ALIGNMENTS

Accession	Result 1
LOCUS	AI956440
DEFINITION	AI956440 527 bp mRNA EST 20-AUG-1999
	u177c02.y1 Sugano mouse kidney mkl4 Mus musculus cDNA clone
	IMAGE:216578 5 similar to gb:U14751 Mouse ATP receptor (MOUSE);
	mRNA sequence.

ACCESSION	AI956440
NID	95749149
VERSION	AI956440.1
KEYWORDS	GI:5749149
SOURCE	EST.
ORGANISM	house mouse.
	Mus musculus

REFERENCE	1 (bases 1 to 527)
AUTHORS	Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.

TITLE	The WashU-NCI Mouse EST Project 1999
JOURNAL	Unpublished (1999)
COMMENT	On Jun 5, 1998 this sequence version replaced g1:3189116.

Contact: Maria M/Wasnu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:1001254
Trace considered overall poor quality
Seq primer: custom primer used
High quality sequence stop 1.

FEATURES	Location/Qualifiers
source	1..527
	/organism="Mus musculus"
	/strain="C57BL"
	/db_xref="taxon:10090"
	/clone_image:213578"
	/clone_lib="Sugano mouse kidney mk1a"
	/sex="female"
	/dev_stage="adult"

```

/lab_host="DHIOB"
/note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
(CACCTGTGtg); Site_2: DraIII (CACCATGTg); 1st strand cDNA
was primed with an oligo(dt) primer
[Atgtgcctttttttttttttttt]; double-stranded cDNA was
ligated to a DraIII adaptor [Tctggccttccttg], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGg, 3' site CACCATGTg). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCGTCTCAAAAGCGGC and 3' end
primer CGACCTGCGAGCTGACACA."
BASE COUNT      108 a      147 c      141 g      130 t      1 others
ORIGIN

```

Query Match 9.2%; Score 130.8; DB 63; Length 527;
Best Local Similarity 63.0%; Pred. No. 2.3e-23;
Matches 218; Conservative 0; Mismatches 127; Indels 1; Gaps 1;

Db 61 GAGCTGCCATCAAGTGTAGATTAAAGCAGAGACTTCAAGTACATGCTGCAGCCCATGTGC 120

Db 121 TATGGTGAAGTCCGGGTGCTGCICGAGTTGTGCTGAACTCGAGGGTCTTTATATGTTCTTA 180

Db 181 TGGCGACATTAACCTGGAACGCTTACACCACTACATGTTTCACCTGGCAGATTAGAC 240

Db 241 TCCTCTACGACAGCTCCCTGCGCTTGTATTACTACGCCCGATTGACCACTGG 300

[illegible]

Db 361 AGCATCCCTCTTCTCCTACGCTGAT-AGCGTCACGGGTGCCCTGAGTA 405

RESULT	2
FR0019937	
LOCUS	586 bp DNA
DEFINITION	FR0019937
DESCRIPTION	F1.rubridus GSS sequence, clone 037P03ac9, genomic survey sequence.
FEATURES	
1	10-DEC-1997
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

KEYWORDS
VERSION
AL018224.1 GI:2679192
GSS: genome survey sequence.
SOURCE
Fugu rubripes.

Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodonidae; Fugu

NOTES	
TITLE	Williams, G., Clark, M., Smith, S., Meek, S., Warner, S., Umranna, Y., Williams, G., and Brenner, S.
Direct Submission	

COMMENT	Vector: bluescript II KS	EMAIL: DIONEI@PENG.MPG.MPG.AC.UK
	V-type: phagemid	
	nature: no	

One pass dye-terminator sequencing of cosmid cloned genomic sequence.

FEATURES

Location/Qualifiers

[illegible]

FEATURES	SOURCE
Location/Qualifiers	1. 316
/organism="Homo sapiens"	
/db_xref="ATCC (Inhost):194234"	
/db_xref="taxon:9606"	
/clone_1ib="umbilical vein endothelial cells II"	
/cell_type="endothelial cell"	
/dev_stage="fetus"	
/note="Organ: umbilical vein; Vector: pBluescript SK-;	
Site_1: EcoRI; Site_2: XhoI"	
67 a 106 c 72 g 71 t	

	Query Match	Similarity	5.2%	Score 74.2	DB 31	Length 316
	Best Local	Similarity	59.1%	Pred. No. 5.8e-09		
	Matches 127	Conservative	0	Mismatches 89	Indels 0	Gaps 0
QY	383	ATGCACGGCCACACATGTTCCACCTGGCATTGTGCACACACCTTGTATGTGCTGTGGC	442			
Db	48	ATGACACGCACTGTACACTCTTGAACCTGGCATTGCCACCTGTGGTGTCTCCACCA	107			
QY	443	TGCCACACCCATCTACTACTTATATGACGCCCAACACCATGGGCCCTTGGCAGTAGATCT	502			
Db	108	TCCCAAGTCTGGTGGTCACTCTCTGTGCACCAACCAAGTGGCCCATGGGCGAGCTTCACT	167			

OY		GCAGATTCGCGGGTTCTTCTTATTGTGAACCCACTCACTGTCCTTCTCACC	562
Dδ		GCAAAGTCACACACCTCATCTTCTCCATCAACTCTTCGGCACATTTTCTCTCAGT	227
OY		GATCACCAGTCACCGCTACTTGCGCATCTGCCAC	597
Dδ		GCAATGACGTGACCGCTACTCTTCCTACCTAAC	262

RESULT	7
LOCUS	AA331167
DEFINITION	AA331167 261 bp mRNA EST 21-APR-1997
ACCESSION	U5135211 Embryo, 8 week 1 Homo sapiens cDNA 5' end similar to vasocactive intestinal peptide receptor RDC1, mRNA sequence.
NID	AA331167
VERSION	91983628
KEYWORDS	AA331167.1 GI:1983628
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
	1 (bases 1 to 261)
	Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,

TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL	Nature 377 (6547 Suppl.), 3-174 (1995)
MEDLINE	96026280
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1407040. Other ESTs: EST179908 THC105954
	Center, Vancouver, B.C.

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/cdb/hgi.hgi.html>)
Seq primer: M3 Reverse.
Location/Qualifiers
1..261
source

BASE COUNT	59 a	90 c	54 g	56 t	2 others
ORIGIN					
Query Match		5.28;	Score 73.8;	DB 32;	Length 261
Best Local Similarity		58.68;	Pred. No. 6.7e-09;		


```

/db_xref="taxon:9606"
/clone="DKFZ585M0824"
/clone_lib="586 (synonym: hutel)"
/tissue_type="uterus"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Vector: pSport1; Site-1: Not; Site-2: SalI/MluI"
BASE COUNT      172 a      106 c      112 g      199 t      3 others
ORIGIN

```

Query Match	5.0%;	Score 70.8;	DB 64;	Length 592;
Best Local Similarity	53.7%;	Pred. No. 5.9e-08;		
Matches 189;	Conservative 0;	Mismatches 159;	Indels 4;	Gaps 2

OY	257	ACTGTGGTTGATGAGGATTTCACAGTTCACCTCGTGCTGTAGCATGACATTGTCT	316
Db	229	ACTGGTCTTAATAAAGACTCCCTTTAAAGAACACTTTGTATNNGGTGCATGTTGACATGGTGT	288
OY	317	TTTGTCGTGGCCTTAGGGCCTTAAGCCCCAACCCCTATGCGTCTTCATCTTCCGCCCTCGAC	376
Db	289	TTTGCTGTGGTCTAATATNCACAAATGTGTGGCAATACATTTTCATTCGTGGCTCTCAAG	348
OY	377	CCTGGAGTSCAAGCGCACCCACCTACATNTTCCACCCT-GGCATTGTCAGACACCTGTATGTG	435
Db	349	TCCGAATAAGAAGACACACTTACATGTATTAACCTGGGCAANGTAGACTGCTTTTGTGT	408
OY	436	CTGTGCGTCCACCACCTCATCTACTATTATATGACAGCCACAACCACTGGCCCTTTGGCACT	495
Db	409	TTTACTTTACCCCTTCAGAGATTTTTTACTCTCACAACAGGAAT---TGGCATTTTGGAGAT	465
OY	496	GAGATCTSCAAGTTCGTGCCGCTTTCTTTTCTATTGGAACCTCTACTGCAAGTGTCTTTTC	555
Db	466	TTACTTTGTAGATTTCTGTGATCTGTTTATACCAAATGTATGCGAAGAACATTTGTTTC	525
OY	556	CTCAGCTSCATCAGAGGCGACAGCGTACTGGGCGATCTGCCAACCCACTCTGGG	607
Db	526	TTTAACCTTATATAGTGTATGATCGATTTCTGGCAATTTGTCAACCAATTTAAAGG	577

RESULT	10
LOCUS	AA177828
DEFINITION	AA177828 529 bp mRNA EST 16-FEB-1997 mt0f02.r1 Soares mouse 3NBDS Mus musculus cDNA clone IMAGE:620379 5' similar to TR:G663021 G663021 PROTEININSE ACTIVATED RECEPTOR 2. ; mRNA sequence.
ACCESSION	AA177828
NID	g1759090
VERSION	AA177828.1 GI:1759090
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Crustacea; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 529)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellensberg,K., Steptoe,M., Tan,F., Underwood,R., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterson,R.
TITLE	The Mashu-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced g1:1393684.

Contact: Maira M/Mouse EST Project
WashU-HIMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LINL; contact the
IMAG Consortium (info@image.lnl.gov) for further information.

MGI:381203	
Seq primer: -28M13 rev2 from Amersham	
High quality sequence stop: 494.	
Location/Qualifiers	
1..529	
FEATURES	
SOURCE	

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="IMAGE:620379"
/clone_1lb="Soares mouse 3Nbms"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a self strand cDNA
polylinker. Site_1: Not I. Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - o1a5o(Gr) primer [5].
TGTTACCAATCTGAAGTGAGGCGGCGCCCTCTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and-Eco RI sites of the modified p773 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."

```

Query Match	4.98;	Score 70;	DB 29;	Length 529;
Best Local Similarity	50.38;	Pred. NO. 8.9e-08;		
Matches 200; Conservative	0;	Mismatches 195;	Indels 3;	Gaps 1

[illegible]

RESULT	11
AA611127	
LOCUS	AA611127 650 bp mRNA EST 01-OCT-1987
DEFINITION	vogt108.r1 Soares mouse mammary gland NBMG Mus musculus cDNA clone IMAGE:1054407 5' similar to gb:D13814 TYPE-1A ANGIOTENSIN II RECEPTOR (HUMAN); , mRNA sequence.
ACCESSION	AA611127
NID	G2461325
VERSION	AA611127.1 GI:2461325
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

Db	11	ATCTGCGCCCTGGCTGACTTAATGCTTTTGGTGGACATTGGCCCTGTGGCGTGTCTATACCG	70
Oy	467	CAGCCCAACACCACTGTGGCCCTTTGGCACTGAATCTGCAATTCGTCCGCTTCTTTC	528
Db	71	CTATGGAATATACCGCTGGCGCCCTTCCGCAATCACTATGAAATATGCTTTCGGCCAGCCTCA	130
Oy	537	ATTGGACCTCTACAGCAGTGCCTTTTCCACACTGATAGCGTGACCGCACTGG	586
Db	131	GTTTAACTCTACGCCACGCGTGTTCCTGCTCAGGTGTCTCAAGATGATGCTACTGG	190
Oy	587	GCATTCGCCACCACTTGGGCACTACGCTGGGGCGCGCTCGGCTCGCAGGCTTCT	646
Db	191	CCATGTGTCACCCCGATGAAAGTCTCGCTCCGCCGACAGATCTGTGGCCAAATGACCT	250
Oy	647	GCCCTGACGTTTGGTGGATCGTAAGCCGGCTGCTCTGTCGCCAACCTGTC	696
Db	231	GCATATCATCTGCGTATGCTGCTGCTTTGGCCAGTTTCCAGCCGTATC	300

RESULT 13
AA205847
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA205847 816 bp mRNA EST 12-MAR-1998
zg90c02.1 Stratiagene neuroepithelium (9377231) Homo sapiens CDNA
clone IMAGE:645026 5' similar to gb:M64749 G PROTEIN-COUPLED
RECEPTOR RDC1 HOMOLOG (HUMAN);, mRNA sequence.
AA205847
AA205847 g1801218
AA205847.1 GI:1801218
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 816)
Hillier, L., Allen, M., Bowles, L., Dubnue, T., Geisel, G., Jost, S.,
Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Thaisang, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
On Jan 24, 1995 this sequence version replaced gi:1643356.

FEATURES

Location/Qualifiers
1. .816

CDNA Library Preparation: M.
M.D., Ph.D.

Bento Soares, Ph.D.

3

BASE COUNT	181 a	238 c	193 g	197 t	7 others
ORIGIN					
Query Match		4.64;	Score 65.8;	DB 30;	Length 816

	Best Local Similarity	52.3%	Pred. No.	1.3e-06	Mismatches	155	Indels	1	Gaps	1
Oy	278	TCAAGTTCATCTGCTGCCTGTGAACATTGACAGTTGCTTTTGCTGGCGCTTGCA	337							
Dd	237	TGCTCTCTACACGCGTCCCTTCATTACATTTTATCTCTGTCATCGCAGTAATTGCCA	296							
Oy	338	ACGCCCCAACCCCATATGCGTCTTCATCTTTCGCCCTCCGACCCTGGGATSCAAGCGCACCT	397							
Dd	297	ACTCCGTGGTGCTGGGTGAATAATCCAGGCCAAGACCAAGGCTATGACAGCACTGCT	356							
Oy	338	ACAATTTCCACCTGGCATTTGACAGACCTGTATGTGCTGTGCTGCCACCTTCATCT	457							
Dd	357	ACATCTTGAACTCGGCACATTCCGCCAGCTGTGGGTGTCTCTACCATCCAGTCTGGGTG	416							
Oy	438	ACTATTATGACGCCACACACACATGSGCCCTTTGGAGCTGAGATCTGCAGATTGCTCGCT	517							
Dd	417	TCAGTCTGTGACAGCACACACAGTGCGCCCATGGGAAACCTCACGTCGAAGAATCACACACC	476							
Oy	518	TTCTTTTATTTGGAAGCTC-TACTGCAAGTGTCCCTTTTCTCACCCTGATCAGCTGCAC	576							
Dd	477	TCATCTTTCTCATCAACCTCTTGTGGGAGCAATTTTCTTCTCACGCTGATGAGCGTGCAC	536							
Oy	577	CGCTACCTGGGCATCTGCCAC	597							
Dd	537	CCTAACCTTCATCACTTAC	557							

RESULT	14
LOCUS	AI739295/c
DEFINITION	AI739295 570 bp mRNA EST 18-JUN-1999
DESCRIPTION	w150b03.x1 NCI-CGAP-Co16 Homo sapiens CDNA clone IMAGE:2391725 3' similar to TR:076067 076067 PROTEASE-ACTIVATED RECEPTOR 4., mRNA sequence.
ACCESSION	AI739295
NID	G5101276
VERSION	AI739295.1 GI:5101276
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
	Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 570)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3189094.

Contact: Robert Strausberg, Ph.D.
 Tel.: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck,
 M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/URL at:
www-bio.lnl.gov/bdip/image/image.html
 Seq primer: -40UP from Gldco
 High quality sequence stop: 453.
 Location/Qualifiers
 1..570
FEATURES
 source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2391725"
/clone_1d="NCI_CGAP_Col6"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pTRAD-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI;

```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 04:51:36 ; Search time 104.56 Seconds
(without alignments)
3419.324 Million cell updates/sec

Title: US-09-077-173a-1

Perfect score: 1429
Sequence: 1 AAGCAGCTTGCGTAGGGC.....GCCACGAGCAGGATATC 1429

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Word size: 0

Number of hits that pass the threshold: 623170

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1424.2	99.7	1429	1	T74321	Human P2Y4 receptor
2	350.4	24.5	1842	1	O88134	Human P20 receptor
3	218.6	15.3	984	1	T18368	Human placenta G-p
4	210.8	14.8	984	1	T18367	Mouse pancreas G-p
5	210.6	14.7	984	1	T47730	Human placenta pur
6	210.6	14.7	984	1	T70000	P2U putative rec
7	128.4	9.0	1901	1	O66178	Seven transmembran
8	128.4	9.0	2453	1	T44092	Human G-protein th
9	128.4	9.0	1901	1	V18356	Human R12 seven tr
10	107.6	7.5	1679	1	O37107	New platelet facco
11	107.6	7.5	1679	1	O80522	Human lymphocyte p
12	107.6	7.5	1679	1	O99009	Chemokine superfam
13	104	7.3	2427	1	X07369	Human P2Y11 recept
14	98.6	6.9	1477	1	O84557	Murine C140 recept
15	98.6	6.9	2732	1	O84557	Murine C140 recept
16	98.6	6.9	1477	1	T32036	Murine C140 recept
17	98.6	6.9	2732	1	T32038	Murine C140 recept
18	96.4	6.7	1670	1	V26557	Human IP-10/MY re
19	94.8	6.6	1876	1	T72800	Human G-protein ch
20	94.6	6.6	2216	1	O66556	Murine delta opioi
21	93.8	6.5	1435	1	X06948	HEB4D54 cDNA parti
22	93	6.5	1821	1	O56700	Sequence of murine
23	93	6.5	2272	1	O75927	Mouse delta opioi
24	93	6.5	2218	1	V49253	Mouse delta opiate
25	92.2	6.5	1597	1	V07250	DNA encoding a G-p
26	91.2	6.4	1425	1	O84558	Human C140 recept
27	91.2	6.4	1414	1	O84560	Human C140 recept
28	91.2	6.4	1425	1	T32037	Human C140 recept
29	91.2	6.4	1594	1	T32039	Human C140 recept
30	91.2	6.4	1594	1	X06949	Human C140 recept
31	90.8	6.4	1872	1	O66176	Seven transmembran
32	90.8	6.4	1872	1	V18354	Human R20 seven tr
33	89	6.2	1301	1	T33905	Human G-protein co
34	85.2	5.9	1296	1	O45657	Human somatostatin
35	84.2	5.9	1312	1	T93367	Mouse thrombin rec
36	83.4	5.8	1361	1	X16671	G-protein coupled
37	81.4	5.7	2864	1	V40374	Human G-protein co
38	81.4	5.7	1395	1	V41400	Human 7-transmembr
39	80.6	5.6	1110	1	O29155	Pituitary somatost

40	80.2	5.6	1080	1	T30385	Probe based on ang
41	79	5.5	1200	1	O30011	Sequence encoding
42	79	5.5	1200	1	O29949	Recombinant high a
43	79	5.5	1029	1	T68663	Pig growth hormone
44	79	5.5	1029	1	T69755	Swine growth hormo
45	78.8	5.5	3480	1	O28568	Human thrombin rec

ALIGNMENTS

RESULT	1	
ID	T74321	
AC	T74321 standard; DNA; 1429 BP.	
DT	31-MAR-1998 (first entry)	
DE	Human P2Y4 receptor genomic DNA.	
KW	Receptor; P2Y4; pyrimidine binding; uridine triphosphate; UTP; ss.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	CDS	181..1278
FT		/*tag="a
FT		/product="P2Y4 receptor
PN	MO9719170-A1.	
PD	29-MAY-1997.	
PF	21-NOV-1996; BE0123.	
PR	21-NOV-1995; EP-870124.	
PA	(EURO-) EUROSCREEN SA.	
PI	Boeynaems J, Commun D, Parmentier M, Piroton S;	
DR	WPI: 97-402177/37.	
DR	P-PSDB: W23606.	
PT	Receptor having preference for pyrimidine over purine nucleotide(s)	
PT	- especially uridine triphosphate, agonist and antagonists of which	
PS	are useful in treatment of cystic fibrosis	
PS	Claim 14; Figure 1; 56pp; English.	
CC	This sequence encodes a novel human P2 receptor, P2Y4, which has a	
CC	preference for pyrimidine binding, especially uridine triphosphate.	
CC	This receptor could be used to screen for novel drugs which	
CC	specifically bind to it. Transgenic animals could be used to determine	
CC	the physiological effects of expressing varying levels of the receptor	
CC	or to identify novel agonists or antagonists. The agonists and	
CC	antagonists of human P2Y4 may be used, e.g., in treatment of cystic	
CC	fibrosis.	
SQ	Sequence 1429 BP; 242 A; 451 C; 379 G; 357 T;	
Query Match		
Best Local Similarity 99.7%; Score 1424.2; DB 1; Length 1429;		
Matches 1426; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
OY	1 AAGGAGCTTGGGTGGGAGCAGCTAGCTAGAGCAACCCAGATGCGTGTGACGT	60
DB	1 AAGGAGCTTGGGTGGGAGCAGCTAGCTAGAGCAACCCAGATGCGTGTGACGT	60
OY	61 CTCCTAGTCTTCAACACAGCTCTCCCTGCTACTTTTCTGCTCAGCTAGGAT	120
DB	61 CTCCTAGTCTTCAACACAGCTCTCCCTGCTACTTTTCTGCTCAGCTAGGAT	120
OY	121 GGGGTTGGGCAAGAAATCTGCAACCTCACTTCTCCCTTCCATCTCCAGGGGGCC	180
DB	121 GGGGTTGGGCAAGAAATCTGCAACCTCACTTCTCCCTTCCATCTCCAGGGGGCC	180
OY	181 ATGGCAGTACAGAGTCTCCCTGTTGAGATCCCTAGGCTTCCAGGCTTGGCAGC	240
DB	181 ATGGCAGTACAGAGTCTCCCTGTTGAGATCCCTAGGCTTCCAGGCTTGGCAGC	240
OY	241 AGTGAAGTGAAGTGAAGTGTGATGAGATTCAGATTATCTGCTGCTG	300
DB	241 AGTGAAGTGAAGTGAAGTGTGATGAGATTCAGATTATCTGCTGCTG	300
OY	301 AGCTATGAGTGTCTTGTGCTGGGCTTAAAGCCCAACCTATGCTCTTC	360
DB	301 AGCTATGAGTGTCTTGTGCTGGGCTTAAAGCCCAACCTATGCTCTTC	360

OY	361	ATCTTCGGCCCTCCGACACCTCGGGAATGCAAGCGGCACCTACATGTTTCCACTCGGCAATTGTCA	420
Db	361	ATCTTCGGCCCTCCGACACCTCGGGAATGCAAGCGGCACCTACATGTTTCCACTCGGCAATTGTCA	420
OY	421	GACACCTGTATGTGCTGTGCTCTCCCAACCTCATCTACATTAATGACGCCAACAACAC	480
Db	421	GACACCTGTATGTGCTGTGCTCTCCCAACCTCATCTACATTAATGACGCCAACAACAC	480
OY	481	TGGCCCTTTGGCACTGAGATCTGCAGATTGCTCGCGCTTTCTTTTATTTGAACCTCTAC	540
Db	481	TGGCCCTTTGGCACTGAGATCTGCAGATTGCTCGCGCTTTCTTTTATTTGAACCTCTAC	540
OY	541	TGCAGTGTCTTTTCTTCACCTGCATAGCGTGCACCGCTACTCTGGGCATCTGCCACCA	600
Db	541	TGCAGTGTCTTTTCTTCACCTGCATAGCGTGCACCGCTACTCTGGGCATCTGCCACCA	600
OY	601	CTTGCGGCATACGCTGSGGGGCGCGCTCGGCTGCAAGCGCTTCTGCGCTTGCGCACTTTGG	660
Db	601	CTTGCGGCATACGCTGSGGGGCGCGCTCGGCTGCAAGCGCTTCTGCGCTTGCGCACTTTGG	660
OY	661	TTGGTCGTAGCCGGGCTGCTGTGTCGCCCAACCTTTCTTTGTCAACACGACAACAAAGG	720
Db	661	TTGGTCGTAGCCGGGCTGCTGTGTCGCCCAACCTTTCTTTGTCAACACGACAACAAAGG	720
OY	721	ACCAACCTCTCTGTGCCATGACACCACCTGCGCTGGAAGATTGACCATATGTGCATTC	780
Db	721	ACCAACCTCTCTGTGCCATGACACCACCTGCGCTGGAAGATTGACCATATGTGCATTC	780
OY	781	AGCTCGGGGGCATGSGGGGCGCTTTTGGGSGTCCCTGCTGGTACCTTTGTGGTAT	840
Db	781	AGCTCGGGGGCATGSGGGGCGCTTTTGGGSGTCCCTGCTGGTACCTTTGTGGTAT	840
OY	841	GGACTCATGSGCTCTGCTGCGCTGTATCAAGCCTTGTCAGAGCCTGCACAGCGTCTTGCG	900
Db	841	GGACTCATGSGCTCTGCTGCGCTGTATCAAGCCTTGTCAGAGCCTGCACAGCGTCTTGCG	900
OY	901	CTCGGCTCTCTCGGCACCACTAGCTGTGTCGCTACTTTGCTGTGCTTGTGCTGCT	960
Db	901	CTCGGCTCTCTCGGCACCACTAGCTGTGTCGCTACTTTGCTGTGCTTGTGCTGCT	960
OY	961	TTTCACACATACCCGACACCTTTACTACTCTGGCCAGGCTGTGGAGCTGATGCCGAGTA	1020
Db	961	TTTCACACATACCCGACACCTTTACTACTCTGGCCAGGCTGTGGAGCTGATGCCGAGTA	1020
OY	1021	CTGAACATTTGTCAACGTCGTCTAATAAGTATGATCGGCGCCCTGGCCAGTCCCAACGCTGC	1080
Db	1021	CTGAACATTTGTCAACGTCGTCTAATAAGTATGATCGGCGCCCTGGCCAGTCCCAACGCTGC	1080
OY	1081	CTGATCCTGTGCTCTACTTTGCTCAGCTGGGAGAAATATGCAAGCTGATCCGTACGTC	1140
Db	1081	CTGATCCTGTGCTCTACTTTGCTCAGCTGGGAGAAATATGCAAGCTGATCCGTACGTC	1140
OY	1141	TGTGTGTGTGGCAAGCCCCACGCCGACGCGCTCTTCCCTGGCACTAGTGTCCCTG	1200
Db	1141	TGTGTGTGTGGCAAGCCCCACGCCGACGCGCTCTTCCCTGGCACTAGTGTCCCTG	1200
OY	1201	CCTGAGATATGACACTCTGAGTGGGGGGCCACCCCCAGAGACATGACTGCTCTACTCT	1260
Db	1201	CCTGAGATATGACACTCTGAGTGGGGGGCCACCCCCAGAGACATGACTGCTCTACTCT	1260
OY	1261	AGGCGATATGATCTTAACACGGGAAGCCGGCAAGTAGAGAAAGGGGATGAGTACAGG	1320
Db	1261	AGGCGATATGATCTTAACACGGGAAGCCGGCAAGTAGAGAAAGGGGATGAGTACAGG	1320
OY	1321	GCAAGAGTGAAGGAACCCAAATAGTATACCTGTGTAAGTGTCTTCTCTTTTCCAGAC	1380
Db	1321	GCAAGAGTGAAGGAACCCAAATAGTATACCTGTGTAAGTGTCTTCTCTTTTCCAGAC	1380
OY	1381	TCTGGAGAGAAAGCCCTCAACCTTAGAGGTTTGCCACGAGCAGGAGATATC 1429	
Db	1381	TCTGGAGAGAAAGCCCTCAACCTTAGAGGTTTGCCACGAGCAGGAGATATC 1429	

RESULT	2
ID	Q08134 standard; cDNA; 1842 BP.
AC	Q08134;
DT	29-NOV-1995 (first entry)
DE	Human P20 receptor gene.
KW	Epithelial mucosa; mucus; cystic fibrosis; asthma; ss;
OS	chronic bronchitis.
TX	Homo sapiens.
FT	Key
FT	Location/Qualifiers
FT	57..1181
FT	/*tag= a
PN	W09510538-A.
PD	20-APR-1995.
PE	04-OCT-1994; U11260.
PR	15-OCT-1993; U5-138137.
PPA	(UMOR) UNIV MISSOURI.
PPA	(UNMC-) UNIV NORTH CAROLINA.
PI	Boucher RC, Erd Lf, Harden TK, Lustig KD, Parr CE;
PI	Sullivan DM, Turner JT, Weisman GA;
DR	WPI; 95-169967/22.
DR	P-PSDB: R72457.
PT	DNA encoding human P20 receptor and null cells expressing the
PT	receptors for stimulating or inhibiting growth of cultures of
PT	mammalian cells; and for treating diseases of airway epithelial
PT	mucosa, e.g. asthma.
PS	Claim 3; Page 34; 47pp; English.
CC	The sequence is that of the human P20 receptor gene. The gene may
CC	be used to treat diseases of epithelial mucosal surfaces by
CC	enhancing or inhibiting mucus prodn. The gene may be used to treat
CC	cystic fibrosis; asthma and chronic bronchitis.
CC	See also Q08135-40.
SO	Sequence 1842 BP; 364 A; 557 C; 528 G; 393 T;

Query Match	24.5%	Score 350.4	DB 1	Length 1842
Best Local Similarity	64.9%	Pred. No. 2.9e-87		
Matches 570	Conservative 0	Mismatches 296	Indels 12	Gaps 3
QY	243	TGAGGTGGAGAGCTGGAGCTGTGGTTGGATGATGAGATTTCACAGTTCATCTCTGCTGGTGGAG	302	
Db	113	TGAGCTGGGCTTCACAGTCCGCTTTACAGGAGACTTGAATACGTGCTCTCCCTGTGTC	172	
QY	303	CTATGCACTGTCTTTTGTCTGGGCTTGGGCTTAAAGCCGCCAACCCCTATGGCTCTTCAT	362	
Db	173	CTAGCGCGGAGTGTGCGTGGCTTGGGCTGTGTGAACGCCGTGGGCTCTACATCTTCTT	232	
QY	363	CTTCCGCTCCGACCCCTGGATGACAGGCCACCTACATGTTCCACTGGCATTTGACA	422	
Db	233	GTGCGGCTTCAAGACCTGGAAAGCTGCCACCATATATATGTTCACACTGGCTGTGTGCA	292	
QY	423	CACCTGATATGCTGTGGCTGGCCGCCACCCCTACTTACTATTTATGAGCCACACCACTG	482	
Db	293	TGCACTGATGATGGGCTCCCTCCCTGCTGCTTACTTACTACAGCCGGGGAGCACACG	352	
QY	483	GCCCTTTGGACATGATCTGCAATGTCTGCCGCTTTCTTTCTATTTGGAACTCTACG	542	
Db	353	GCCCTTAGCAGAGGTGCTCTGCAAGCTGTGGGCTTCTCTTCTACACCAACCTTTACG	412	
QY	543	CAGTGTCTTTTCTTCACCTGCATCAGGATGACAGCGCTACCTGGGACATCTGCCACCACT	602	
Db	413	CAGCACTCTTCTTCACCTGCATCAGGATGACAGCGGTGTCTGGGCGTCTTACGACCTCT	472	
QY	603	TGCGGCATAGGCTTGGGGCCGCTCGCTCGCAGGCTTCTCTGCTGGCAGTTTGGT	662	
Db	473	GCGCTCCCTGGCGTGGGGCCGCGGCCGCTACAGCTCGCGGGGTGGCGCGGTGTGGAT	532	
QY	663	GTCGTACACGGCTCCCTGCTGTCGCCCACTGTTTGTACACACACCAACAAGGAGC	722	
Db	533	GTTGGTGTGGCTGTGCCAGGCGCCCGGTCTTACTTTGTACACACCAACAGCGCGCGG--	591	
QY	723	CACGCTCTGTGCCATGACACCACTCGGCTGAGAGTTTGACCACTATGTGCACTTCAG	782	


```

Db 171 CCGACCCGACGCGCGGTGACACCCCTAACCTGTCTGCTGACCTGCTATATGCTG 230
Qy 438 GTGCGCCGACCCCTATCTACTATATGACAGCCACACCACTGGCCCTTGGCACTGA 497
Db 231 CTCCCTGCGCCCTGCTATCTATCACTATGCGCAAGGTGATCGGCGCTTGGCACTT 230
Qy 498 GATCTGCAAGTTGCTGCGCTTCTTTCTATTTGGAACCTTACTGACGCTGCTTTCCT 557
Db 291 CGCTGCGCGCGCTGCTGCGCTTCTTCTTATGCGCAACCTGACAGGGGAGATCTCTCT 350
Qy 558 CACCTGCAATGAGGTGACAGCGCTACCTGCGCATCTGCGACCACTGCGGCACTACGC-- 616
Db 351 CACCTGCAATGAGGTGACAGCGCTACCTGCGCATCTGCGACCACTGCGGCACTACGC-- 616
Qy 616 -TGGGGCGCGCGCTGCGCTGCGAGCGCTTCTGCTGCGGAGTTGTTGTTGCTAGCGG 674
Db 411 ACCTGGGGCGCGCGCTGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 470
Qy 675 CTGCGCTGCGCGCGCAACCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 734
Db 471 CCAAGTCTGCGCGCGCAACCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 530
Qy 735 CCATGACACCACTGCGCTGAGAGAGTTTGACACATATGCTACTTCAAGCTGGCGGTAT 794
Db 531 TTATGACCTGACGCGCGCGCTGCGCTGCGCGCAACCTATATGCTTATGATGCTCTAC 590
Qy 795 GGGGCGCTCTTGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 854
Db 591 TGTCAATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 650
Qy 855 TGGCGCTGATACGCC-----CTTGGCAGGCTGACACAGTCTTCTGCTGCTGCTGCTG 908
Db 651 CCGCTGCTGCGCGCGCAAGAGTGGCGCGCGCAAGCTGCTGCGCGCAAGCGGCTGCGCA 710
Qy 909 TCTCCGACACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 968
Db 711 GCGCGCGCAAGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 770
Qy 968 CACCGCGCAACCTTAC---TACCTGCGCAAGGCTGCTGCAAGCTGACGCTGCGAGTGA 1025
Db 771 CACCAAGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 830
Qy 1025 CATTTGCAACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1085
Db 831 GCGCTTGGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 890
Qy 1086 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 891 CCCCATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 945

```

```

RESULT 6
T70000
ID T70000 standard; cDNA; 984 bp.
AC T70000.1998 (first entry)
DE P2U purinergic receptor gene homologue.
KW Primer; PCR; extend; partial cDNA; human cathepsin B; sequence analysis;
KW complete gene; detection; amplification; 5'-untranslated region;
KW promoter; 8S.
OS Homo sapiens.
PN M09638591-A1.
PD 05-DEC-1996.
PF 03-JUN-1996; U08501.
PR 01-DEC-1995; US-566334.
PR 02-JUN-1995; US-459046.
PR 03-JUN-1995; US-462355.
PR 07-JUN-1995; US-487112.
PR 15-NOV-1995; US-006809.
PA (INCY-) INCYTE PHARM INC.
PI Guegler RJ.
WP1: 97-034397/03.

```

```

PT Extending partial cDNA sequences by PCR amplification with outwardly
PT extending primers - allowing rapid isolation of long sequences,
PT including complete genes, and detection or amplification of
PT promoters
PS Example 3; Page 37; 64pp; English.
CC This is the full length cDNA of a novel P2U purinergic receptor gene
CC homologue. It was identified using the inventive method and comparison
CC between the partial sequence found in Incyte clone 179696 (from a
CC placental cDNA library) and the GenBank sequence, accession number
CC RNU09402, a G-protein coupled surface receptor from rat. The method
CC comprises: (a) subjecting nucleic acid from a genomic or cDNA library
CC expected to contain partial cDNA, to reaction with a PCR primer pair,
CC which is able to anneal to opposite strands of the template DNA and to
CC initiate outward DNA synthesis, where one is extended in the sense direction, to
CC generate PCR products; and (b) purifying the products, and identifying
CC extended sequences derived from the partial cDNA or genomic DNA. The
CC method is used to produce much longer DNA (complete genes in some cases)
CC from libraries which have been used for sequencing. It may also be used
CC to extend known genomic sequences for detection and amplification of
CC 5'-untranslated, and/or promoter sequences. Up to 96 genes may be
CC processed at a time, and the extended or complete sequence is obtained
CC within 6-10 days, an approximate 10-fold improvement in productivity.
SQ Sequence 984 bp; 157 A; 360 C; 259 G; 208 T.

```

```

Query Match 14.7% Score 210.6; DB 1; Length 984;
Best local similarity 54.6% Pred. No. 8.4e-49;
Matches 489; Conservative 0; Mismatches 394; Indels 12; Gaps 3;

```

```

Qy 258 CTGTTGCTTGTAGAGATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 317
Db 51 CTGTTGCTTGTAGAGATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 110
Qy 318 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 377
Db 111 GCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 170
Qy 378 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 437
Db 171 CCGACCCGACGCGCGGTGACACCCCTATATGCTGCTGCTGCTGCTGCTGCTGCTG 230
Qy 438 GTGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 497
Db 231 CTCCCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 290
Qy 498 GATCTGCAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 557
Db 291 CGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 350
Qy 558 CACCTGCAATGAGGTGACAGCGCTACCTGCGCATCTGCGACCACTGCGGCACTACGC-- 616
Db 351 CACCTGCAATGAGGTGACAGCGCTACCTGCGCATCTGCGACCACTGCGGCACTACGC-- 616
Qy 616 -TGGGGCGCGCGCTGCGCTGCGAGCGCTTCTGCTGCGAGTTGTTGTTGCTAGCGG 674
Db 411 ACCTGGGGCGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 470
Qy 675 CTGCGCTGCGCGCAACCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 734
Db 471 CCAAGTCTGCGCGCGCAACCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 530
Qy 735 CCATGACACCACTGCGCTGAGAGAGTTTGACACATATGCTACTTCAAGCTGGCGGTAT 794
Db 531 TTATGACCTGACGCGCGCGCTGCGCTGCGCGCAACCTATATGCTTATGATGCTCTAC 590
Qy 795 GGGGCGCTCTTGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 854
Db 591 TGTCAATGCGCTTCTGCTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 650
Qy 855 TGGCGCTGATACGCC-----CTTGGCAGGCTGACACAGTCTTCTGCTGCTGCTGCTG 908
Db 651 CCGCGCTGCTGCGCGCGAGATGCGCGCGCAAGAGCTTGGCGCGCGAGAGCGGCTGCGCA 710

```


PR 06-JUN-1995; WO-007180.
 PA (HOMA-) HUMAN GENOME SCI INC.
 PI Gocayne JD, Li Y, Ruben SM;
 DR WPI: 97-043073/04.
 PR P-PSDB: W07617.
 PT Human G-protein thrombin receptor, HIBB69 - useful to identify
 PT (ant)agonists, for treatment of angina, restenosis, wound healing
 etc.
 PS Claim 1: Fig 1A-D: 58pp; English.
 CC A cDNA clone (T44092) codes for a human 7-transmembrane receptor
 CC (W07617) putatively identified as a G-protein thrombin receptor. It
 CC was discovered in a human infant brain cDNA library. G-protein
 CC thrombin-like receptor polynucleotides can be used to produce
 CC recombinant receptor in host cells (e.g. E. coli, COS-7, Sf9), to
 CC design probes for detecting diseases related to receptor gene
 CC mutation, and in gene therapy to treat haemophilia or promote wound
 CC healing; antisense sequences can be used to treat inflammation,
 CC restenosis, angina, etc.
 SQ Sequence 2453 BP; 494 A; 815 C; 640 G; 504 T;

Query Match 9.0%; Score 128.4; DB 1; Length 2453;
 Best Local Similarity 48.9%; Pred. No. 6e-26;
 Matches 447; Conservative 0; Mismatches 446; Indels 21; Gaps 3;

QY 255 GACATGTTGGTTGATGAGATTTCAGTTCAATCCGCTGCTGAGCTATGAGTTGT 314
 DB 610 GCATGTGCGCCAGAGACGCCACTGAGACATGCTGCTGCTCTTACCTTGTGA 669
 QY 315 CTTTGTGCTGGGCTTGGGCTTAAACGCCCAACCCATGCTCTTCACTTCCGCTCG 374
 DB 670 TTTATCCGCTTAAAGTACTGCGCCCTGCTGCTGCTTTCATCCGAGACACA 729
 QY 375 ACCCTGGATGCAAGGCCACTATGTTCCACTGCTGCTGCTGCTGCTGCTGCT 434
 DB 730 GTCCGGACCCCGCCCAAGTGTCTGATGATGCTGCTGCTGCTGCTGCTGCT 789
 QY 435 GCTGCGGCGCCCAACCTATCTATATATGACGCCCAACCACTGCTGCTGCTGCT 494
 DB 790 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 849
 QY 495 TGAGATCTGCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 554
 DB 850 AATGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 909
 QY 555 CTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 614
 DB 910 CTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 969
 QY 615 CTGGGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 674
 DB 970 GCTCGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1029
 QY 675 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 734
 DB 1030 GGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1089
 QY 735 CCATGACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 794
 DB 1090 CCGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1137
 QY 795 GGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 854
 DB 1138 GCGAGTGGCTTCACTTCCGCTTATCAGACAGGCTGCTGCTGCTGCTGCTGCT 1197
 QY 855 TCGCTGATACAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 914
 DB 1198 CAGCTGCGGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1251
 QY 915 CACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 974
 DB 1252 CATGATGCCATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1311

QY 975 CACCATTTACTACCTGCGCCAGGCTGTTGAGAGTACTGCTCCGAG--TACTGACATTGT 1031
 DB 1312 CTCCTGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1371
 QY 1032 CAACGTGCTATTAAGTACTGCGCCCTGCGCAGTGCACACAGTGTGCTGCTGT 1091
 DB 1372 GCGCTGCGAAGCCGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1431
 QY 1092 GCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1151
 DB 1432 CAGTATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1491
 QY 1152 CAGCCCGCAGCCCG 1165
 DB 1492 AAGCTTCAGAGGCC 1505

RESULT 9
 ID V18356 standard; DNA: 1901 BP.
 AC V18356;
 DT 25-SEP-1998 (first entry)
 DE Human R12 seven transmembrane (7TM) receptor genomic DNA.
 KW V28; placenta; seven transmembrane receptor; 7TM; signal transduction;
 KM immunology; inflammation; R12; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 701..1720
 FT /tag- a
 FT /product- "Human R12 seven transmembrane receptor"
 PN US5759804-A.
 PD 02-JUN-1998.
 PF 17-NOV-1993; 153848.
 PR 17-NOV-1992; US-977452.
 PA (ICOS-) ICOS CORP.
 PI Goelska R, Gray PW, Schweickart VL;
 DR WPI: 98-332132/29.
 P-PSDB: W48733.
 PT DNA encoding V28 seven transmembrane receptor polypeptide - useful
 PT for producing recombinant polypeptide and anti-V28 antibodies, and
 PT in screening assays for V28 agonists and antagonists
 PS Example 10; Columns 85-88; 56pp; English.
 CC The present novel sequence represents the human R12 genomic DNA isolated
 CC from a human genomic fetal liver DNA library. The invention claims for
 CC the full length V28 genomic DNA sequence (V18343) isolated from a human
 CC placenta genomic library. The V28 (W48722) and R12 proteins are seven
 CC transmembrane (7TM) receptors which are probably involved in signal
 CC transduction. The invention also claims that cells transformed with
 CC V28 DNA can be used to produce the recombinant polypeptide, to produce
 CC anti-V28 antibodies or in screening assays for V28 agonists or
 CC antagonists. The antibodies, agonists and antagonists could then be
 CC used to modulate V28 receptor-ligand binding, for e.g. in immunological
 CC and/or inflammatory events in vivo.
 SQ Sequence 1901 BP; 381 A; 640 C; 508 G; 372 T;

Query Match 9.0%; Score 128.4; DB 1; Length 1901;
 Best Local Similarity 48.9%; Pred. No. 5.4e-26;
 Matches 447; Conservative 0; Mismatches 446; Indels 21; Gaps 3;

QY 255 GACATGTTGGTTGATGAGATTTCAGTTCAATCCGCTGCTGAGCTATGAGTTGT 314
 DB 763 GCATGTGCGCCAGAGAGAGCCACTGAGAACATGCTGCTGCTGCTGCTGCTGCT 822
 QY 315 CTTTGTGCTGGGCTTGGGCTTAAACGCCCAACCCATGAGCTCTTCACTTCCGCTCG 374
 DB 823 TTTATCCTGCTTAAAGTATGCAATCCCTGCTGCTGCTGCTGCTGCTGCTGCT 883
 QY 375 ACCCTGGATGCAAGGCCACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 934
 DB 883 GTCGGGACCCCGCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 985
 QY 435 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 985


```

Db 943 GCTGGTCTCCCAACCCGCTGGTCTCCCACTTCTGGGAGACACTGGCCATTGGGGA 1002
Qy 495 TGAGATCTGCAAGTCTGCTCCGCTTCTTTCTATGGAACCTCTACTGAGTCTTTT 554
Db 1003 AATCGATGCGGCTGATCCGCGCTTCTCTTCTACTACCATGATGAGCCAGCATCTT 1062
Qy 555 CTTACCTGATCAGCTGCAACCGCTTACCTGGGATGTGCCACCCACTTGGGGCAAC 614
Db 1063 CTTACCTGATCAGCTGCAACCGCTTCTCTTCTGGGATGTGCCACCCACTTGGGCA 1122
Qy 615 CTGGGAGCGGCTGGGCTGGGAGGCTTCTCTGCTGGGAGTGTGGTGGTGGGAGG 674
Db 1123 GCTCCGAGGCGGCTTACGACACCTGGGCTGCTCTCTGAGGAGGAGGAGGCT 1182
Qy 675 CTGCTGCTGCTCCCAACCTGCTTCTTCTACACACGACCAAGAGGAGGAGGAGG 734
Db 1183 GGGCATGGGCGCGCTGCTGGAGCCACAGACCGTGGACAGACACAGAGGAGGCTG 1242
Qy 735 CCAATGACACACCTGGGCTGGAAGATTGACCACTATGTGCACTTCACTGAGCTCAT 794
Db 1243 CCTGACACTTACCGGGAGAGAGGCTCCACCA-----TGCCCTGGTGTCCCT 1290
Qy 795 GGGGCTCTCTTTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 854
Db 1291 GGGAGTGGCTTACCTTCCCTCCCTTATCACCACGCTGACCTGCTGATCATCG 1350
Qy 855 TCGCCTGATACGCTTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 914
Db 1351 CAGCTTCCGCGAGGAGGCTGGG-----TGTGGAGAGAGGAGGCTTCAAGAGAGAG 1404
Qy 915 CACCAATGCTGTGGTGTGCTGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 974
Db 1405 CATATGCGCCTAGTGGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1464
Qy 975 CACCATTTACTACTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 1031
Db 1465 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1534
Qy 1032 CACGCTGCTATTAAGTACTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 1091
Db 1525 GGGCCTGCGCAACCGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1584
Qy 1092 GCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1151
Db 1585 CATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1644
Qy 1152 CAGGCGCCAGGCGCC 1165
Db 1645 AAGGCTCAAGGCGC 1658

```

RESULT 10

```

Q37107
ID Q37107 standard; DNA: 1679 bp.
AC Q37107;
DE 12-MAR-1993 (first entry)
DE New platelet factor 4 receptor superfamily member PF4R11.
KW IL-8R; G-protein coupled receptor family; rhodopsin superfamily;
KW pro-inflammatory cytokine; 8tr; 9; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 369..1487
FT /tag= a
FT /product= PF4R11

```

```

DR P-PSDB; R27793.
PT Isolated human platelet factor 4 super-family receptor
PT dipeptide and corresp. antibodies and DNA - useful as
PT diagnostic and screening agents, and for treating inflammation or
PT PF4R-mediated disorders
PS Claim 7; Fig 5; 78pp; English.
CC The IL-8 receptor cDNA sequence was isolated (see Q29505) and a
CC 874bp sub-fragment of the coding sequence was used as a probe to
CC screen human cell line HL60 and human peripheral blood lymphocyte
CC cDNA libraries. Two new gene sequences were found that are clearly
CC related to the IL-8 receptor. One of these was contained in clone
CC 8tr.9 and is predicted to encode an amino acid sequence which is
CC 36% and 38% identical with the high and low affinity IL-8 receptor
CC sequences, respectively. See also Q37107.
SQ Sequence 1679 bp; 327 A; 532 C; 457 G; 363 T;

Query Match 7.5%; Score 107.6; DB 1; Length 1679;
Best Local Similarity 47.0%; Pred. No. 2.8e-20;
Matches 409; Conservative 0; Mismatches 449; Indels 12; Gaps 2;

Qy 277 TTCAAGTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 336
Db 513 TTCAAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 572
Qy 337 AACGCGCCCAACCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 396
Db 573 AACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 632
Qy 397 TACATGTTCCACCTGGCATTTGTGACAGACCTTGTATGCTGCTGCTGCTGCTGCTGCT 456
Db 633 TTCTGTTTCCACCTGGCGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 692
Qy 457 TACTATTATGAGCCCAACACACTGCGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCT 516
Db 693 GCGGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 746
Qy 517 TTTCTTTTCTATTGGAACCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 576
Db 747 GCGCTGCAAAAGTCAATTTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 806
Qy 577 GCTACCTGGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 636
Db 807 GCTACCTGGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 866
Qy 637 GGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 696
Db 867 CACATCACTGTGGGACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 926
Qy 697 TTTTCAACACGACCAACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 756
Db 927 TTGCGCAAAAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 986
Qy 757 GAGTTTGAACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 816
Db 987 GAGAAACAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1046
Qy 817 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 876
Db 1047 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1106
Qy 877 GAGCTGACAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 936
Db 1107 CAGGCGCCAGGCGGCGCTGACGCGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1166
Qy 937 GCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 996
Db 1167 ATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1226
Qy 997 CTGTTGGAAGCTGAC-----TGCCGAGTACTGAACTGTGCAAGTGTCTATTAAGTG 1050
Db 1227 CTGAAGCGCGTGGCAATATCTGCAAGCTGAATAGCTCTCTCCCGTGGGCAATCAGATG 1286

```


1051 ACTCGGCCCCGCGCCAGTGCACACAGCTGCTGATCTGTGCTTACTTCTACTGGG 1110
1287 TGTGAGTTCTCTGGCGCTGGCCACATGCTGCTCAACCCATGCTTACACTTTCGCCGGC 1346
1111 GACAAATATGAGTGCAGTCCGCTCAGCTC 1140
1347 GTGAAGTTCCGCAATGACTGTGCGGGCTC 1376

RESULT 11
080522
ID 080522 standard; cDNA; 1679 BP.
AC 080522; (first entry)
DT 18-JUN-1995
DE Human lymphocyte PF4AR cDNA.
KW Interleukin-8 receptor; IL-8 receptor; PF4AR;
KW platelet factor superfamily; receptor; lymphocyte; chemotactic;
KW inflammation; inflammatory disease; arthritis; emphysema; cystic;
KW fibrosis; colitis; bronchitis; meningitis; therapeutic; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 369..1487
FT /tag- a

W09428931-A.
PD 22-DEC-1994.
PE 07-JUN-1994; 006380.
PR 11-JUN-1993; US-076093.
PA (GENTH) GENE TECH INC.
PI Chuntharapai A, Hebert C, Kim KJ, Lee J;
DR WPI: 95-036114/05.
DR P-PSDB: R68813.
PT Treatment of inflammatory disorders - by administering an
PT antibody capable of binding a platelet factor 4 superfamily
PT receptor polypeptide
PS Disclosure: Page 56-58; 83pp; English.
CC 2 PF4AR members were identified by probing lambda libraries from
CC human monocytic-like cell line HL-60 and human peripheral blood
CC lymphocytes using a large fragment of IL-8 receptor DNA (full
CC sequence given in Q80520). The nucleotide sequences of the 2
CC PF4ARs are given in Q80521 and Q80522 and their respective
CC amino acid sequences in R68812 and R68813.
SO Sequence 1679 BP; 327 A; 532 C; 457 G; 363 T;

Query Match 7.5%; Score 107.6; DB 1; Length 1679;
Best Local Similarity 47.0%; Pred. No. 2.8e-20;
Matches 409; Conservative 0; Mismatches 449; Indels 12; Gaps 2;

277 TTCAAGTTCATCTGCTGCTGCTGAGACATGATGAGTCTTGTGCTGGGCTTGGCCTT 336
513 TTCAAGGCGCTGTGCTGCGCGCTGACGCTATCTTCTCTGGGCGTGATCGGC 572
337 AAGCGCCCAACCTATGCTCTTTCATCTTCGCGCTTCGACCTGGGATGCAAGCGGCACC 396
573 AAGCTCTGCTGCTGCTGCTGCTGAGCGGACCGGACAGACAGGATTCACAGGAGACC 632
397 TACATGTTCCAGCTGCGATGTCAGACAGCTGTATGCTGTGCTGCTGCCACCTCATC 456
633 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 692
457 TACATTTATGAGCCACACACACCTGCTTGGCACTGAGATGCAAGTTGCTGCCG 516
693 GCGGAGGCTCTGTGGG-----CTGGGCTGCGGGAGCTTCTCTGCAAACTGTGATT 746
517 TTCTTTTCTATGGAACCTCTACTGACATGCTTCTTCTGCTGCTGCTGCTGCTGCTG 576
747 GCGCTGACAAAGTAACTTACTGACAGAGCTCTCTGCTGCTGCTGCTGCTGCTGCTG 806
577 GCGTACTGCGGATCTGCGACCACTCTGCGGACATGAGCTGCGGCGCGCTGCTGCTGCA 636
807 GCGTACTGCGGATCTGCGACCACTCTGCGGACATGAGCTGCGGCGCGCTGCTGCTGCA 866
637 GCGTACTGCGGATCTGCGACCACTCTGCGGACATGAGCTGCGGCGCGCTGCTGCTGCTG 696

867 CACATCACCCTGTGGGACCATCTGGCTGTGGGCTTCTCTTGGCTTGGCAGAGATTCTC 926
697 TTTGTCACACAGCAGACAAAGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 756
927 TTGGCCAAAGTAAAGGACGACATCAACAACTCCCTGCTGCTGCTGCTGCTGCTGCTG 986
757 GAGTTTGAACCTATGATGACATTCAGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTG 816
987 GAGAACCAAGCAGAAAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1046
817 TCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 876
1047 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1106
877 GCGTCTGACACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 936
1107 CAGGCCACAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1166
937 GCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 996
1167 ATCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1226
997 CTGTTGGAAGCTGAC-----TGCCGAGTACTGAACTGTCAAGTGTCTATAAAGTG 1050
1227 CTGAAGGCGGCTGACATACATGCTGCAAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 1286
1051 ACTCGGCCCCGCGCCAGTGCACACAGCTGCTGATCTGTGCTTACTTCTACTGCTGCGG 1110
1287 TGTGAGTTCTCTGGCGCTGGCCACATGCTGCTCAACCCATGCTCACTTTCGCCGGC 1346
1111 GACAAATATGAGTGCAGTCCGCTCAGCTC 1140
1347 GTGAAGTTCCGCAATGACTGTGCGGGCTC 1376

RESULT 12
099009
ID 099009 standard; cDNA; 1679 BP.
AC 099009;
DT 26-MAR-1996 (first entry)
DE Chemokine superfamily receptor coding sequence.
KW Interleukin; IL-8; inflammation; psoriasis; dermatitis;
KW rheumatoid arthritis; inflammatory bowel disease;
KW chronic lung inflammation; treatment; antibody;
KW affinity purification; detection; ss.
OS Homo sapiens.
PN US5440021-A.
PD 08-AUG-1995.
PE 29-MAR-1991; 677211.
PR 29-MAR-1991; US-677211.
PR 25-FEB-1994; US-202056.
PA (CHUN) CHUNTHARAPAI A.
PA (HEBE) HEBERT C.
PA (KIMK) KIM K J.
PA (LEEJ) LEE J.
PI Chuntharapai A, Hebert C, Kim KJ, Lee J;
DR WPI: 95-283151/37.
DR P-PSDB: R92239.
PT New antibodies against interleukin 8 type B receptor - used to treat
PT or prevent inflammation, also for detecting receptor expression and
PT purification.
PS Example 2; Columns 51-54; 62pp; English.
CC Antibodies directed against the interleukin-8 receptor B can be used
CC to treat or prevent inflammation e.g. psoriasis, dermatitis,
CC rheumatoid arthritis and particularly inflammatory bowel disease and
CC chronic lung inflammation. When immobilised, these antibodies may
CC be used to detect interleukin-8 receptor B expression in cells and
CC tissues and for affinity purification of interleukin-8 receptor B
CC from cells. This sequence is an additional chemokine superfamily
CC receptor which was identified by probing lambda libraries of genomic
CC DNA from a human monocytic-like cell line (L-60) and human peripheral
CC blood lymphocytes using a large fragment of the interleukin-8 type

Db 1708 GCCCGAAGCCTGCGACCCAGACAGCCGCGCTGAGCGCTGCCGCTGGCTCTG 1767
Oy 667 GTAGCCGGCTGCTGCGCCCAACCTGTCTT 698
Db 1768 GCCGCGCTGCTGCGCAGCCACACTGACTT 1799

RESULT 14

084557 084557 standard: DNA; 1477 BP.
AC 084557: 22-AUG-1995 (first entry)
DE Murine C140 receptor genomic DNA.
KM G-protein-coupled receptor; G-protein; C140 receptor; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 232..1188
ET /*tag= a
PN WO9503318-A.
PD 02-FEB-1995.
PE 26-JUL-1994; U08536.
PR 26-JUL-1993; US-097938.
PS (COR-) COR THERAPEUTICS.
PI Scarborough RM, Sundelin J;
DR WPI: 95-075182/10.
DR P-PSDB: R66920.
PT New DNA encoding recombinant C140 receptor - and novel agonists
PT and antagonists and specific antibodies with therapeutic and
PT diagnostic applications.
PS Disclosure; Fig 1; 57P; English.
CC A mouse cosmid genomic library (obtd. from Dr R.A. Wetzel, Washington
CC Univ. School of Medicine, St Louis, Missouri) was screened with two
CC 32P-labeled oligos corresp. to bp 190-249 and 742-801 of the bovine
CC substance K receptor cDNA. In one of the clones isolated (C140) the
CC hybridizing region was localised to a 3.7 kb psti fragment. This
CC fragment was subcloned into pBluescript vector. The hybridising and
CC adjacent regions were sequenced. The nt sequence and the deduced AA
CC sequence are given in Q84557 & R69920 respectively. The published
CC DNA sequence appears to have been printed with the last two lines
CC in the wrong order. This has been corrected in this Geneset entry.
SQ Sequence 1477 BP; 341 A; 414 C; 320 G; 402 T;

Query Match 6.9%; Score 98.6; DB 1; Length 1477;
Best Local Similarity 57.2%; Pred. No. 8.1e-18;
Matches 179; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Oy 287 TCCTGCTGCTGAGCTATGAGTGTCTTGTGCTGGGCTTGGGCTTACGCCCA 346
Db 452 TCCTTCTTCGGGCGTGTACATTAATGTGTGATGGTTGCCAGTAATGGCATGG 511
Oy 347 CCTATGAGCTTCTCATCTTCGCTCCGACCTGGATGAGACGGCCACCTACATGTTCC 406
Db 512 CCTCTGAGATCTTCTTTCGGAACGAAGAAGAACACCCCGCGATTTTACATGSCCA 571
Oy 407 ACCTGCAATGTAGACACCTTGTATGTGTGCTGCTGCCACCCATCATTAATATG 466
Db 572 ACCTGCTTGGCGGACCTCTCTCTGTATGTGTTCCCTGAAATGCTCTTACACACC 631
Oy 467 CAGGCCAACAACCTGGCCCTTGGCACTGAGATCTGCAAGTTGCTCCGCTTTTCT 526
Db 632 TACATGGCAACAACCTGGCTACGAGGAGGCCCTGTCAAGTGTCTATGGCTTTTCT 691
Oy 527 ATTGAACCTCTACTGAGTGTCTTCTTCTACCTGATGAGGCTGACCGCTACCTGG 586
Db 692 ATGGTAACAGTATGTCTCATCTCTTCAATGACTGCTCCTCAGCGTCAAGAGTACTGG 751
Oy 587 GCATCTGCCACC 599
Db 752 TGATCGTGAACCC 764

RESULT 15

084559 084559 standard: cDNA; 2732 BP.

AC 084559: 22-AUG-1995 (first entry)
DE Murine C140 receptor cDNA.
KM G-protein-coupled receptor; G-protein; C140 receptor; ss.
OS Mus musculus.

FH Key Location/Qualifiers
FT cds 73..1272
ET /*tag= a

PN WO9503318-A.
PD 02-FEB-1995.
PE 26-JUL-1994; U08536.
PR 26-JUL-1993; US-097938.
PS (COR-) COR THERAPEUTICS.
PI Scarborough RM, Sundelin J;
DR WPI: 95-075182/10.
DR P-PSDB: R66922.
PT New DNA encoding recombinant C140 receptor - and novel agonists
PT and antagonists and specific antibodies with therapeutic and
PT diagnostic applications.
PS Claim 1; Fig 10; 57P; English.
CC A cDNA library from a mouse stomach was constructed in lambda gt10
CC and screened with a probe encompassing the C140 genomic clone (see
CC Q84557). A single phage clone was isolated and cut with EcoRI. The
CC insert was cloned in pBluescript and pSG5 and sequenced. The
CC complete nt sequence and deduced AA sequence is given in Q84559 &
CC R66992. 5' RACE resulted in the addition of only 27 bps to the 5'
CC end. The 5' end of the apparent coding region differs from the 5'
CC end of the ORF of genomic DNA; it is believed that the 5' end of
CC the cDNA sequence is correct.
SQ Sequence 2732 BP; 650 A; 676 C; 633 G; 773 T;

Query Match 6.9%; Score 98.6; DB 1; Length 2732;
Best Local Similarity 57.2%; Pred. No. 1e-17;
Matches 179; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Oy 287 TCCTGCTGCTGAGCTATGAGTGTCTTGTGCTGGGCTTGGGCTTACGCCCA 346
Db 305 TCCTTCTTCGGGCGTGTACATTAATGTGTGATGGTTGCCAGTAATGGCATGG 364
Oy 347 CCTATGAGCTTCTCATCTTCGCTCCGACCTGGATGAGACGGCCACCTACATGTTCC 406
Db 365 CCTCTGAGATCTTCTTTCGGAACGAAGAAGAACACCCCGCGATTTTACATGSCCA 424
Oy 407 ACCTGCAATGTAGACACCTTGTATGTGTGCTGCTGCCACCCATCATTAATATG 466
Db 425 ACCTGCTTGGCGGACCTCTCTGTGTATGTGTTCCCTGAAATGCTCTTACACACC 484
Oy 467 CAGGCCAACAACCTGGCCCTTGGCACTGAGATCTGCAAGTTGCTCCGCTTTTCT 526
Db 485 TACATGGCAACAACCTGGCTACGAGGAGGCCCTGTGCAAGTGTCTATGGCTTTTCT 544
Oy 527 ATTGAACCTCTACTGAGTGTCTTCTTCTACCTGATGAGGCTGACCGCTACCTGG 586
Db 545 ATGGTAACAGTATGTCTCATCTCTTCAATGACTGCTCCTCAGCGTCAAGAGTACTGG 604
Oy 587 GCATCTGCCACC 599
Db 605 TGATCGTGAACCC 617

Search completed: December 29, 1999, 07:20:25
Job time: 8929 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 1999, 04:48:33 ; Search time 601.47 Seconds

(without alignments)
10158.894 Million cell updates/sec

Title: US-09-077-173a-1

Perfect score: 1429

Sequence: 1 AAGGAGCTGTGGGTAGGGGC.....GCCACGAGGACAGGATATC 1429

Scoring table: IDENTITY_NUC

Searched: 780561 seqs, 2137953050 residues

Database : GenBank:*

Word size : 0

Number of hits that pass the threshold : 1561122

1: gb_dal:*
2: gb_dal:*
3: gb_dal:*
4: gb_dal:*
5: gb_dal:*
6: gb_dal:*
7: gb_dal:*
8: gb_dal:*
9: gb_dal:*
10: gb_dal:*
11: gb_dal:*
12: gb_dal:*
13: gb_dal:*
14: gb_dal:*
15: gb_dal:*
16: gb_dal:*
17: gb_dal:*
18: gb_dal:*
19: gb_dal:*
20: gb_dal:*
21: gb_dal:*
22: gb_dal:*
23: gb_dal:*
24: gb_dal:*
25: gb_dal:*
26: gb_dal:*
27: gb_dal:*
28: gb_dal:*
29: gb_dal:*
30: gb_dal:*
31: gb_dal:*
32: gb_dal:*
33: gb_dal:*
34: gb_dal:*
35: gb_dal:*
36: gb_dal:*
37: gb_dal:*
38: gb_dal:*
39: gb_dal:*
40: gb_dal:*
41: gb_dal:*
42: gb_dal:*
43: gb_dal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1424.2	99.7	1429	5	A62985	A62985 Sequence 1
2	1424.2	99.7	1429	9	HSP2Y4	X91852 H.sapiens p
3	1363.6	95.4	1651	11	HSU40223	U40223 Human urid1
4	1228.8	86.0	1232	9	HSCPCRE	X96557 H.sapiens g
5	802.2	56.1	2684	12	RNP2Y4RG	Y14705 Rattus norv
6	794.4	55.6	1177	12	RNP2Y4REC	Y11433 R.norvegicu
7	416.4	29.1	1977	4	AF031897	AF031897 Meleagris
8	379.2	26.5	2025	11	HSU07225	U07225 Human P2U n
9	379.2	26.5	2025	11	S74902	S74902 HP2U-P2U nu
10	359.2	25.1	2011	12	RAMP2U	L46665 Rattus norv
11	359.2	25.1	1668	12	RNP56839	U56839 Rattus norv
12	357.6	25.0	1918	12	RNP09402	U09402 Rattus norv
13	352	24.6	1842	5	I34573	I34573 Sequence 1
14	352	24.6	1842	5	I36524	I36524 Sequence 1
15	352	24.6	1842	5	I76280	I76280 Sequence 1
16	346.6	24.3	1131	12	S83099	S83099 P2U recepto
17	343.4	24.0	2138	12	MUSP20REC	L14751 Mouse ATP r
18	326.8	22.9	2361	4	XLPEY8	X99953 X.laevigis mr
19	218.6	15.3	984	5	E12487	E12487 Human CDNA
20	218.6	15.3	1571	9	HSP2Y6	X97058 H.sapiens m
21	218.6	15.3	1507	10	HSU52464	U52464 Human P2 pu
22	218.6	15.3	1539	11	AF007891	AF007891 Homo sapi
23	218.6	15.3	1832	11	AF007892	AF007892 Homo sapi
24	210.8	14.8	984	5	E12485	E12485 Mouse CDNA
25	209.2	14.6	2352	11	AF007893	AF007893 Homo sapi
26	208.2	14.6	1922	12	RATNGPCP2R	D63665 Rat mRNA fo
27	195.8	13.7	1163	4	GDP2Y3	X98283 G.domesticu
28	195.4	13.7	1498	4	GDATP2REC	X73268 G.domesticu
29	187.8	13.1	1074	4	AF069555	AF069555 Meleagris
30	185.8	13.0	1192	4	AF012103	AF012103 Meleagris
31	185.8	13.0	1180	4	MGU09842	U09842 Meleagris g
32	167.4	11.7	1666	3	BTP2YREC	X87628 B.taurus mr
33	167.4	11.7	1308	3	BTG34041	U34041 Bos taurus
34	147.4	10.3	16178	9	HSAS569	AJ006945 Human P2Y
35	147.4	10.3	1426	10	HSU42029	U42029 Human P2Y1
36	147.4	10.3	2424	11	HSU42030	U42030 Human P2Y1
37	147.4	10.3	1312	11	S81950	S81950 P2 purinose
38	141.2	9.9	3204	12	RNU22830	U22830 Rattus norv
39	138	9.7	400	3	AF005153	AF005153 Bos tauru
40	136.2	9.5	1293	12	MMU22829	U22829 Mus muscu
41	134.6	9.4	3055	9	HSATPRMR	Z49205 H.sapiens m
42	134.6	9.4	12630	12	MMU245636	AJ245636 Mus muscu
43	134.6	9.4	3055	13	G28604	G28604 human STS S
44	128.4	9.0	2070	9	HSP2Y6G	Y12546 H.sapiens m
45	128.4	9.0	288945	42	AC010878	AC010878 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS A62985 1429 bp DNA
DEFINITION Sequence 1 from Patent WO9719170.
ACCESSION A62985
NID 93716857
VERSION A62985.1 GI:3716857
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1429)
AUTHORS Commun,D., Piroton,S., Parmentier,M. and Boeynaems,J.
TITLE RECEPTOR AND NUCLEIC ACID MOLECULE ENCODING SAID RECEPTOR
JOURNAL Patent: WO 9719170-A 1 29-MAY-1997;
FEATURES
source Location/Qualifiers
1..1429
/organism="unidentified"

gene	CDS	242 a	451 c	379 g	357 t
gene					
CDS					
	/organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="lambda Charon Iva" 181..1278 /gene="p2y4" 181..1278 /gene="p2y4" /codon_start=1 /product="uridine nucleotide receptor" /protein_id="CAA62963.1" /db_xref="PID:e204061" /db_xref="PID:g1124905" /db_xref="GI:1124905" /translation="MASPRESSILRSLGSPGSGSEVLDCEPDEDFKFIILPSTAV VFVAGLGINAPTLMTIFIRLRKPDATATYHMLALSDTLVLSLPTLIYYAANHP FGEELCKFVRLFEYWNLCISVLFICISVHRYLGICHLRALRWGRPLAGLICLAVW LVVAGCLVPLNLEFVYTSNKGTYVLCDDTTRDEEDHYVFSVAWGLLEGPCLVTLV CGYLARLYOPLPGSAOSSRLRSLRIIVALLVYFAVCPFHRTIYYALALEA DCRVNLINYYVYKTRPLASANSCLDPVLYLLTGDKIRQRLQLCGGGKPOPRTAASS LALVSLPDDSSCRWATPQDSSCSTPRADRL"				

Query Match	99.7%	Score 1424.2;	DB 9;	Length 1429;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 1426;	Conservative	0;	Mismatches 3;	Indels 0;
			Gaps	0;

QY	1	AAGGAGAGCTGGGGTGGGGGGCCAGGGAGCCGAGTGGACCCAGTGGCGTTGCTAGCT	60
	1		
Db	1	AAGGAGAGCTGGGGTGGGGGGCCAGGGAGCCGAGTGGAGTGGACCCAGTGGCGTTGCTAGCT	60
QY	61	CTCCCTAGTGCCTTCACCACTGCTCTCCCTGCTACTCTTTTGTCTCCAGCTCAGGGAT	120
Db	61	CTCCCTAGTGCCTTCACCACTGCTCTCCCTGCTACTCTTTTGTCTCCAGCTCAGGGAT	120
QY	121	GGGGGTGGGCGAAGGAAATCCCTGCCACCTCACTTCTCCCTTCCCATCTCCAGGGGGCC	180
Db	121	GGGGGTGGGCGAAGGAAATCCCTGCCACCTCACTTCTCCCTTCCCATCTCCAGGGGGCC	180
QY	181	ATGGCAGTACAGAGTCTCCCTGTTGTGATCCCTAGGCGCCAGCCCAAGGTCCTGGCAGC	240
Db	181	ATGGCAGTACAGAGTCTCCCTGTTGTGATCCCTAGGCGCCAGCCCAAGGTCCTGGCAGC	240
QY	241	AGTGAAGTGGAGAGCTGGAGCTGTTGGTTTGATGAGGATTTCAAGTTCATCTGCTGCTGTG	300
Db	241	AGTGAAGTGGAGAGCTGGAGCTGTTGGTTTGATGAGGATTTCAAGTTCATCTGCTGCTGTG	300
QY	301	AGCTATGAGATGTTGTTGTGCTGGGGCTTGGGGCTTAAAGCGCCCAACCTATAGGCTTTC	360
Db	301	AGCTATGAGATGTTGTTGTGCTGGGGCTTGGGGCTTAAAGCGCCCAACCTATAGGCTTTC	360
QY	361	ATCTTCGCCCTCCGACCCCTGGAGATGCACAGGCGCACACATGATGTCACAGTGGATTGCA	420
Db	361	ATCTTCGCCCTCCGACCCCTGGAGATGCACAGGCGCACACATGATGTCACAGTGGATTGCA	420
QY	421	GACACCTGTATGTGCTGTGCGTGCACCCTCATCTATTAATTGACAGCCCAACCAAC	480
Db	421	GACACCTGTATGTGCTGTGCGTGCACCCTCATCTATTAATTGACAGCCCAACCAAC	480
QY	481	TGGCCCTTGGAGACGAGATGTGCAAGTTCGTCGGCTTCTTTTCTATTTGGAACCTTAC	540
Db	481	TGGCCCTTGGAGACGAGATGTGCAAGTTCGTCGGCTTCTTTTCTATTTGGAACCTTAC	540
QY	541	TGCAGTGTCTTTTCTCAGCTGCATGCAGCTGCACGCTACAGGAGCATCTGCACACCA	600
Db	541	TGCAGTGTCTTTTCTCAGCTGCATGCAGCTGCACGCTACAGGAGCATCTGCACACCA	600
QY	601	CTTGGGCACTACGCTGGGGGGCGCCCTGCTCGAGGCGCTTCTCTGCTGGCAGTTTG	660
Db	601	CTTGGGCACTACGCTGGGGGGCGCCCTGCTCGAGGCGCTTCTCTGCTGGCAGTTTG	660

QY	661	TTTGTCGTAAGCCGGCTCCCTCGTGGCCCAACCTCTTCTTGTGTACAAACGACAAACAAGG	720
Db	661	TTTGGTCGTAAGCCGGCTCCCTCGTGGCCCAACCTCTTCTTGTGTACAAACGACAAACAAGG	720
QY	721	ACCAACCGCTCGTGTCCATGACACCACTCGGCGCTGAAGAGATTGACACATGTGCACCTTC	780
Db	721	ACCAACCGCTCGTGTCCATGACACCACTCGGCGCTGAAGAGATTGACACATGTGCACCTTC	780
QY	781	AGCTCGGGGTCATGGGGCTGCTCTTGGCGTGCCTGCGTGCCTACTTGTGTTGCTAT	840
Db	781	AGCTCGGGGTCATGGGGCTGCTCTTGGCGTGCCTGCGTGCCTACTTGTGTTGCTAT	840
QY	841	GGACTCATAGGTGCTGCGCTGTATACGCTTGTCCAGGCTGTGCACAGTGTCTTCTGCG	900
Db	841	GGACTCATAGGTGCTGCGCTGTATACGCTTGTCCAGGCTGTGCACAGTGTCTTCTGCG	900
QY	901	CTCCGCTCTTCGCCACCATAGCTGTGTGTCTACAGTCTTTGCTGTCTGCTCGGCT	960
Db	901	CTCCGCTCTTCGCCACCATAGCTGTGTGTGTCTACAGTCTTTGCTGTCTGCTCGGCT	960
QY	961	TTTCCATACACCCGACACATTACTACTGTCGGCGGCTGTTGGAACTGACTGCGGAGTA	1020
Db	961	TTTCCATACACCCGACACATTACTACTGTCGGCGGCTGTTGGAACTGACTGCGGAGTA	1020
QY	1021	CTGAACATTGTCAACGCTGTCTATAAAGTACGTGCGCCCTGGCGAGTCCAAAGCTGC	1080
Db	1021	CTGAACATTGTCAACGCTGTCTATAAAGTACGTGCGCCCTGGCGAGTCCAAAGCTGC	1080
QY	1081	CTGGATCCTGTGCTCTACTGTCACAGTGGGGGCAAAATATGAGTCACTGCTGACGTC	1140
Db	1081	CTGGATCCTGTGCTCTACTGTCACAGTGGGGGCAAAATATGAGTCACTGCTGACGTC	1140
QY	1141	TGTGGTGTGTGCAAGCCCGACGCGCGACGCTGCTTCCCTGTGCACTAGTGTCCCTG	1200
Db	1141	TGTGGTGTGTGCAAGCCCGACGCGCGACGCTGCTTCCCTGTGCACTAGTGTCCCTG	1200
QY	1201	CCTGAGGATAGCAGCTCAGGTGGGGCGGACCCGCCAGAGACATGATGCTCTACTCTT	1260
Db	1201	CCTGAGGATAGCAGCTCAGGTGGGGCGGACCCGCCAGAGACATGATGCTCTACTCTT	1260
QY	1261	AGGGCAGATATATCTTCAACACGGGAAGCCGCGCAAGTGAAGAAAAAGGGATATAGTGCAG	1320
Db	1261	AGGGCAGATATATCTTCAACACGGGAAGCCGCGCAAGTGAAGAAAAAGGGATATAGTGCAG	1320
QY	1321	GCAGAGGTGAGGAAGAACCAATAGATAGTACTGTGTAAGTGTCTCTTCCCTTTTCCAGAGC	1380
Db	1321	GCAGAGGTGAGGAAGAACCAATAGATAGTACTGTGTAAGTGTCTCTTCCCTTTTCCAGAGC	1380
QY	1381	TCTGGAGAGAAAGCCCTCACCTCAGGCTTGCCACGAGGACAGGATATC	1429
Db	1381	TCTGGAGAGAAAGCCCTCACCTCAGGCTTGCCACGAGGACAGGATATC	1429
RESULT 3			
LOCUS	HSU40223	1651 bp	DNA
DEFINITION	Human uridine nucleotide receptor (UNR) gene, complete cds.		
ACCESSION	U40223.1		
VERSION	GI:1117912		
KEYWORDS	G protein-coupled receptor; purinoceptor; PCR; intronless; UTP		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 (bases 1 to 1651)		
JOURNAL	Nguyen,T., Erb,L., Weisman,G.A., Marchese,A., Heng,H.H.,		
MEDLINE	Garard,R.C., George,S.R., Turner,J.T. and O'Dowd,B.F.		
REFERENCE	Cloning, expression, and chromosomal localization of the human		
	uridine nucleotide receptor gene		
	J. Biol. Chem. 270 (52), 30845-30848 (1995)		
	2 (bases 1 to 1651)		

DB	1381	CTTGACCTGCTGCTACTTCTGCTACTGCGGACAA	1116
oy	1081	CTGATCCGTGCTCTACTGCTACTGCTACTGCGGACAA	1116
RESULT	11		
LOCUS	RNU56839	1688 bp	03-FEB-1998
DEFINITION	Rattus norvegicus P2u receptor protein mRNA, complete cds.		
ACCSSION	U56839		
NID	91336124		
VERSION	U56839.1	GI:1336124	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
CDS			
BASE COUNT	357 a	513 c	443 g 375 t
ORIGIN			

Query Match	25.1%	Score 359.2	DB 12	Length 1688
Best Local Similarity	63.9%	Pred. No.7.6e-80		
Matches	560	Conservative	0	Mismatches 313; Indels 3; Gaps 1;
QY	244	GAGGTGAGCGTGGACGTGTGGTGTGATGAGATTTCAAGTTCATCTGCTGGCTGTGAGC	303	
Db	199	GAACTGGGTATCAAAATGCTGCTTCAACAGAGCACTTCAAGTATGTCTGTGGCCGTGCC	258	
OY	304	TATGAGTGTCTTGTGCTGGGCTTTGGGCGCTTAAGCGCCCAACCTTAAGCTCTTATC	363	
Db	259	TATGGCGTGTGTGGCTGTCTGGGCTGTGTCCTGAAGCTGTGGCCCTTACATCTTCTCG	318	
OY	364	TTCCGCTTCGACCTTGGGATGCAAGCGCACCTTACATGTTTGCACCTGGCATTTGTACAG	423	
Db	319	TGCCCTCCACAAAGCTGTGACACCTGCCACACACTTCAAGTTTCACCTGGAGTTTCTGAC	378	
OY	424	ACCTGTATGTCTGTGGTGGCCCAACCTCATCTTATATGAGGCCAACAAACCATGG	483	
Db	379	TCTCTCTACGACGCTCCCTGCGCGTGTGTATTACTAGCCCAAGGTATGCCACTGCG	438	
OY	484	CCCTTTGGCACTGAGATCTGCAGATTGCTGCGCTTTCTTTCTATTGGAACCTTACTGC	543	

[illegible]

CDS 359..1483
 /gene="p2u"
 /codon_start=1
 /product="purinergic receptor"
 /protein_id="AA61565.1"
 /db_xref="PID:9563808"
 /db_xref="GI:563808"
 /translation="MAAGLDSWNSITNGTGEDELGYKRFENEDFKYVLLPVSIGVVC
 VLGLNVALVIFYLCRLKTNWASTIYFHLAVSDSLAASLPLVYVYAGDHMPSS
 TVLKLVPFLFYINLXCSILFLTCSIVSRSLRPLSLHSGARARARVAAVWVIL
 VLACOTPVLYEVTTSVRGTRITGHDTSRELSEFHVAVSYMLGLLAVPEPSITLVY
 VLMARLLKPAVGTGGVLPARKKSRVRIALVAFALCELFPHYTRILYISFRSLDS
 CHITNAIMAKIKTRPLASANSCLDPVIFYTAGRLVRFADAPAPTEIPSPARRK
 LGLHPRNRTDVRDLSDLSIDSRSTESTPAGESETKRL"
 359..1480
 mat_peptide
 gene
 3'UTR
 BASE COUNT 404 a 581 c 512 g 421 t
 ORIGIN
 Query Match 25.0% Score 357.6; DB 12; Length 1918;
 Best Local Similarity 63.8%; Pred. No. 1.9e-79;
 Matches 559; Conservative 0; Mismatches 314; Indels 3; Gaps 1;
 Oy 244 GAGGTGAGAGCTGAGATGTTGGTTGATGAGATTCAAGTTGATCCTGCTGCGTGTAGC 303
 Db 416 GAACGTGGGTAAATAATGTCTGCTTCAACGAGAGCTTCAAGTATGTCTGCTGCGCTTTC 475
 Oy 304 TATGCAAGTTGCTTTGTGTGTGGCTTGGGCTTAAAGCCCAACCTATAGCTTTCATC 363
 Db 476 TATGCGGT 535
 Oy 364 TTCCGCGTCCGAGCCTGGATGATCAAGCGCCACCTTCAATGTTCCACCTGGATTTGAC 423
 Db 536 TCCGCGCTTCAAGAGCTGGAAAGCCTTCCACACCTTCAATGTTCCACCTGGAGTTTCTG 595
 Oy 424 ACCATTATGTCTGTCTGT 483
 Db 596 TCTCTCTACGACGCTCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 655
 Oy 484 CCGTTTGGCAGATGATCTGCAAGTTGCTGCGCTTCTTTCTTATTTGGAACCTTACTGC 543
 Db 656 CCATTAGACAGAGTCTGTGCAAGGTGGTGCTTCTTTCTTTCTTCTTCTTCTTCTTCTTCT 715
 Oy 544 AGTGTCTTTTCTCTCACTGATCAAGCGTGACCGGCTACCTGGGCGATGTGCCACCACTT 603
 Db 716 AGCATCTTCTTCTCTCACTGATCAAGCGTGACCGGAGCGTGGGGGTCTGGGCGCTCTG 775
 Oy 604 CGGGGACTACGCTGGGGGCGCGGCTCGGAGGCTTCTGCTGGCAGATTGGTGG 663
 Db 776 CACTCCCTGCGCTGGGGGCGCATGCGCGCTATGCGCCGCAAGTGGCTGCGGTTGTGGGTG 835
 Oy 664 GTCTGAGCGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 723
 Db 836 CTAGTGTGGGCTGTGCGACAGACCGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 895
 Oy 724 ACCGCTCTGTGCTGATGACACACCACTGCGGCTGAAGATTGACCACTATGTCACTTACG 783
 Db 896 CGAATCACCCTGCAAGACACCTCGGACCGAGAGCTCTTTAGGCAATTTTGTGGCTTACAGC 955
 Oy 784 TCGGGGATCATGGGCTGCTCTTTTGGGCGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 843
 Db 956 TCTGTGATGCTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1015
 Oy 844 CTCATGCGCTGTGCGCTGTATCAGCCCTTGGCAGGCTCTGACAGATGCTCTTTCGCGCTC 903
 Db 1016 CTCATGGCCCGGCTGCTCAAAACCGGCTTATGGGACACAGAGTCTCTCTCGGGCCAG 1075
 Oy 904 CGC---TCTTTCGACCAATAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 960

Db	1076	CGCAAGTCTGTGTCGGACCCATTCGCTTAGTACGTGGCGGCTTGCGCCCTTCGCTTCGCTT	1135
Qy	961	TTCCACATACCCGCGACCATTTACACCTCCGGCCAGGCTGTGGAAGTACTGACTCCGAGTA	1020
Db	1136	TTCCACGTCACCCGAGACCCCTCTATTACTCTTCGCGATCACTTGACCTCAGTTCCACACC	1195
Qy	1021	CTGACATGTGTACAAGTGGTCTATTAAGTAGATCGGCGCCCTGGCCAGTGGCAACAGCTGC	1080
Db	1196	CTCAACGCCCATTAACATGGCGGTATATAGATCACCCGGCCAGTGGCCAGCGCAACAGTTGC	1255
Qy	1081	CTGGATCCCTGTGCTCTACTCTACTCTGCTCACTGGGAGCAAA	1116
Db	1256	CTTGACCCCTGTGCTCTACTCTCTCTGCGAGGCGACAGA	1291
RESULT	13		
134573			
LOCUS	134573	1842 bp	DNA
DEFINITION	Sequence 1 from patent US 5596088.		PAT
ACCESSION	134573		06-FEB-1997
NID	91825364		
VERSION	134573.1	GI:1825364	
KEYWORDS			
SOURCE			
ORGANISM	Unknown.		
REFERENCE	Unknown.		
AUTHORS	Unclassified.		
	1 (bases 1 to 1842)		
	Boucher,R.C., Weisman,G.A., Turner,J.T., Harden,T.K., Parr,C.E.,		
	Sullivan,D.M., Erb,L.J. and Lustig,K.D.		
	DNA Encoding the human P.sub.2U receptor and null cells expressing		
	P.sub.2U receptors		
	Patent: US 5596088-A 1 21-JAN-1997;		
	Location/Qualifiers		
	1..1842		
JOURNAL	/organism="unknown"		
FEATURES			
source			
BASE COUNT	364 a	538 c	527 g 393 t
ORIGIN			

Query Match	Best Local Similarity	24.6%	Score 352;	DB 5;	Length 1842;
Matches 571; Conservative	65.0%;	Pred. No. 4.8e-78;	Mismatches 295;	Indels 12;	Gaps 3
QY 243	TCAGCTGAGAGCTGAGACTGTTGGTTTGATGAGGATTTCAAGTCAATCTGCTGCTGTGAG	302			
Db 113	TGAGCTGGGCTACAGAGTCCGGCTTCACGAGAGACTTCAAGTACGATGCTGCTGCTGTGTC	172			
QY 303	CTATCCAACTTGTCTTTGTGCTGGGCTTGGGCTTAACGCCCAACCCATATGGCTCTTCAAT	362			
Db 173	CTACGGGCTGTGTGCTGCTGGGCTGTGTGTGAACCCCGTGGGCTCTTACATCTTCTT	232			
QY 353	CTTCCGCTTCGACCCTGGATGATGACACGCCACCTACATATGTCACACTGGCATGTGCAGA	422			
Db 233	GTGCGGCTCAAGACCTGGATATGCTGTCCACCATATATATGTTCCACTGTGGCTGTGTCTGA	292			
QY 423	CACCTTATATGTGCTGTGGCTGGCCACCCTCATCTACTATTATGCAAGCCACACCAACTG	482			
Db 293	TGCACTGATATGCGGCTCTCCCTGCTGCTGTCTGTATTACTACAGCCCGGGGACACAG	352			
QY 483	GCCCTTTGGACATGAGATCTGCAGATTGTGCGGCTTTTTCATTTGGAACCTCTACAG	542			
Db 353	GCCCTTCAGCAGGATGCTCTCAAGCTGTGGGCTCTCTCTCTTCAACCAACCTTACAG	412			
QY 543	CAGTGTCTTTTCTCATACCTGATCAGAGGTGCACCGCTACCTGGGCAATCTGCCACCCT	602			
Db 413	CAGCATCTCTTCTCCACACTGATCAGGTGATGACCGGATGTCTGGGCGCTTACGACCTCT	472			
QY 603	TCGGGCACTAGCTGGGGCCGCCCTGCGCTGCGAGGCGCTTCTCTGCTGCGAGTGTGTT	662			
Db 473	GCGCTCCCTGCGCTGGGGCCGGGCCCGCTACGTCGTCGGGGGCGCGGCGCTGTGGGT	532			
QY 663	GGTGTAGCGGGCTGCTGCTGCGCCCAACCTGTCTTTTGTACACACACACACAAAGGAC	722			

Db 533 GTTGGTCTGGCCCTGACGAGCCCGCTGCTACTTGTCAACACACAGCGGGGGG-- 591
QY 723 CACCGTCTGTGGCATACACACACTGCGCCCTGAAGAGTTTACACATATGTGCACTTACG 782
Db 591 -CCGCTAACCTGCCACACACCTGCGGACCGGAGCTTTCACCGCCTTCGTGGCTTACG 649
QY 783 CTGGCGGCTATGGGGGCTGCTTGGGGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 842
Db 650 CTGAGTATGCTGGGCTGCTTCTTGGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 709
QY 843 ACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 896
Db 710 GCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 769
QY 897 TCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 956
Db 770 CAGCGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 829
QY 957 GCGCTTCCATACACCGGCAACCTTACTTACTGCGCGAGCTGCTGCTGCTGCTGCTGCTGCT 1016
Db 830 GCGCTTCCATACACCGGCAACCTTACTTACTGCGCGAGCTGCTGCTGCTGCTGCTGCTGCT 889
QY 1017 AGTACTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1076
Db 890 CACCGTCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 946
QY 1077 CTGCGTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1114
Db 947 TTGCGTTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 984

RESULT 14
LOCUS 136524 1842 bp DNA PAT 13-MAY-1997
DEFINITION Sequence 1 from patent US 5607836.
ACCESSION 136524
NID 92086349
VERSION 136524.1 GI:2086349
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1842)
AUTHORS Boucher,R.C., Weisman,G.A., Turner,J.T., Harden,T.K., Parr,C.E.,
Sullivan,D.M., Erb,L.J. and Lustig,K.D.
TITLE Methods of detecting compounds which bind to the P.sub.2U receptor
JOURNAL Patent: US 5607836-A 1 04-MAR-1997;
FEATURES
source 1.1842
Location/Qualifiers
BASE COUNT 364 a 558 c 527 g 393 t
ORIGIN

Query Match 24.6%; Score 352; DB 5; Length 1842;
Best Local Similarity 65.0%; Pred. No. 4.8e-78;
Matches 571; Conservative 0; Mismatches 295; Indels 12; Gaps 3;

Db 293 TGCATGATGCGGCGCTCCCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 352
QY 483 GCGCTTGGACATGAGATCTGCAAGTTGCTCCGCTTCTTCTTATTTGAACCTTACTG 542
Db 353 GCGCTTGGACATGAGATCTGCAAGTTGCTCCGCTTCTTCTTATTTGAACCTTACTG 412
QY 543 CAGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 602
Db 413 CAGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 472
QY 603 TCGGGCACTAGCTGCGGGGCGCGCGCTGCGCGCGCGCTGCGCGCGCTGCGCGCGCTGCG 662
Db 473 GCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 532
QY 663 GGTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722
Db 533 GTTGGTCTGGCCCTGACGAGCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 591
QY 723 CACCGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 782
Db 591 -CCGCTAACCTGCCACACACCTGCGGACCGGAGCTTTCAGCGCTTTCGAGGCTTACG 649
QY 783 CTGGCGGCTATGGGGGCTGCTTGGGGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 842
Db 650 CTGAGTATGCTGGGCTGCTTCTTGGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 709
QY 843 ACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 896
Db 710 GCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 769
QY 897 TCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 956
Db 770 CAGCGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 829
QY 957 GCGCTTCCATACACCGGCAACCTTACTTACTGCGCGAGCTGCTGCTGCTGCTGCTGCTGCT 1016
Db 830 GCGCTTCCATACACCGGCAACCTTACTTACTGCGCGAGCTGCTGCTGCTGCTGCTGCTGCT 889
QY 1017 AGTACTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1076
Db 890 CACCGTCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 946
QY 1077 CTGCGTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1114
Db 947 TTGCGTTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 984

RESULT 15
LOCUS 176280 1842 bp DNA PAT 03-APR-1998
DEFINITION Sequence 1 from patent US 5691156.
ACCESSION 176280
NID 93012434
VERSION 176280.1 GI:3012434
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1842)
AUTHORS Boucher,R.C., Weisman,G.A., Turner,J.T., Harden,T.K., Parr,C.E.,
Sullivan,D.M., Erb,L.J. and Lustig,K.D.
TITLE Method of inhibiting cell growth with the P.sub.2U receptor
JOURNAL Patent: US 5691156-A 1 25-NOV-1997;
FEATURES
source 1.1842
Location/Qualifiers
BASE COUNT 364 a 558 c 527 g 393 t
ORIGIN

Query Match 24.6%; Score 352; DB 5; Length 1842;
Best Local Similarity 65.0%; Pred. No. 4.8e-78;
Matches 571; Conservative 0; Mismatches 295; Indels 12; Gaps 3;

